

데이터 기반 사회에서 데이터 불평등 개념의 탐색적 논의: 공공과 민간의 데이터 불평등 사례를 중심으로*

서형준**

요약

디지털 전환의 확산으로 인한 데이터 기반 사회의 출현은 데이터에 의한 새로운 불평등을 초래할 수 있다. 하지만 현재까지 데이터 불평등에 대한 논의는 디지털 기기 접근성과 같은 정보격차와 비교하여 부족한 실정이다. 본 연구는 문헌 및 사례 등을 통해 데이터 불평등에 대한 개념 및 현황을 조명하고자 한다. 데이터 불평등을 총 세 가지 유형으로 구분하였으며, 데이터를 수집하고 분석할 수 있는 능력인 데이터 이용능력에 의해 야기되는 불평등, 데이터 활용을 위한 데이터 수집 및 활용 주체의 데이터 운영을 의미하는 데이터 관리에 의해 야기되는 불평등, 데이터를 통한 산출물 등을 창출하는 과정인 데이터 활용에 의해 야기되는 불평등에 따른 데이터 불평등 사례를 제시하였다. 유형별 사례를 통한 시사점으로, 첫째, 데이터 이용능력에 따른 불평등 해소를 위해 데이터 인프라 보급 확대 및 공공기관의 데이터 인식 제고가 요구된다. 둘째, 데이터 관리에 따른 불평등 해소를 위해 데이터 제공자의 권리를 강화해야 한다. 셋째, 데이터 활용에 따른 불평등 해소를 위해 데이터에 대한 과의존성을 줄이고, 다양한 차원의 데이터 확보 노력이 필요하다.

주제어 : 데이터 불평등, 데이터 격차, 데이터 기반 사회, 데이터 기반 행정, 디지털 전환

Exploratory Discussion on Data Inequality in a Data-Driven Society: Focusing on Data Inequality Cases in Public and Private Sectors*

Seo, Hyungjun**

Abstract

A data-driven society due to digital transformation can give rise to new inequality based on data. However, there has not been enough discussion on data inequality compared to digital divide based mainly on access to digital technologies. This study aims to shed light on the concept and current status of data inequality through literature and cases. It classified data inequality into three types: first, inequality of data using results on the ability to collect and analyze data; second, inequality of data management results from the managing body that can collect and utilize data stemming from the data provider; third, inequality of data application result from the process that creates output based on data. According to the three types of data inequality, cases of data inequality in public and private sectors are suggested. As policy implications from cases of data inequality, first, there is a need to expand the supply of data infrastructure and improve the data perception of the public organization for the solution of inequality of data-using ability. Second, the right of the data provider should be strengthened for the solution of inequality of data management. Third, there is a need to reduce the overdependence on data and make an effort to find data in various dimensions for the solution of inequality of data application.

Keywords : data inequality, data divide, data-driven society, data-driven public administration, digital transformation

ReReceived Oct 7, 2024; Revised Nov 1, 2024; Accepted Dec 6, 2024

* This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea(NRF-2022S1A5C2A03093690).

** Senior researcher, The Research Institute for Smart Governance and Policy, Inha University(dongchun01@naver.com, <https://orcid.org/0000-0002-9100-2110>)

I. 서론

4차 산업혁명에 이은 디지털 전환(Digital Transformation)은 2020년 COVID-19 팬데믹을 계기로 비대면을 위시한 사회 전반의 디지털화를 가속화시키고 있다. 디지털 전환은 여타 기술 패러다임과 같이 민간 비즈니스에서 유래된 개념으로, 디지털 전환은 일반적으로 인공지능, 모바일, 사물인터넷, 클라우드 등 디지털 기술을 도입하여 효율성·대응성을 향상시켜 새로운 가치를 창출시키는 조직으로 변모시키는 것을 의미한다(Fitzgerald, et al., 2014; Frankiewicz & Chamorro-Premuzic, 2020; Ismail, et al., 2017; Vial, 2019). 디지털 전환은 기존의 신기술 도입 개념과는 차별화되는 것은 디지털화를 통한 조직의 체질 개선을 수반한다. 이는 궁극적으로 디지털 전환의 영향이 보다 근본적이고 광범위한 변화를 초래함을 의미한다.

디지털 전환의 핵심요소로 D.N.A.(데이터(Data), 네트워크(Network), 인공지능(AI))가 제시되고 있는데, 디지털 기술 활용에 필수적인 데이터의 중요성이 높아지고 있다(Joint Ministries, 2020). Frankiewicz and Chamorro-Premuzic(2020)은 디지털 전환은 단지 디지털 기술에 한정되는 것이 아닌, 특히 데이터 활용 역할을 중요한 요소로 제시하였다. 대표적인 디지털 기술인 인공지능도 결국 가치 있는 데이터를 통해 유의미한 통찰력을 제공한다고 언급하였다. 데이터의 활용은 민간 비즈니스 차원만이 아닌 공공부문의 정책 및 공공서비스에서도 최근 광범위하게 나타나고 있으며, 이러한 경향이 증거 기반 행정(Evidence-based Public Administration)에서 발전한 데이터 기반 행정에 대한 개념이다. OECD는 데이터 기반 행정(Data Driven Public Sector)을 공공부문의 정책, 서비스 설계 및 전달에 있어 데이터를 전략 자산으로 활용하는 것으로, 수요자의 요구 충족과 예측, 더 나은 서비스와 정책 전달, 공공 전반의 데이터 통합·접근·공유·활용을 촉진하기 위해 데이터 거버넌스 구조, 데이터 전달 관련 기제

의 개발을 포함하는 것으로 정의한다(Ubaldi, et al., 2020). 공공부문도 수요자 중심 및 개인맞춤형에 대한 인식이 증대하고, 최근 디지털 플랫폼 정부 및 스마트시티 구현에서도 데이터의 가치는 더욱 높아질 것으로 기대되고 있다.

하지만 이러한 데이터 기반 사회의 편익이 대다수에게 평등하게 돌아갈 수 있다는 기대는 다소 요원하다. 대표적으로 최신기술의 확산에 따라 제기되는 정보격차(Digital Divide)가 대표적이다. 정보격차의 정의는 광범위하나 공통적으로 사회 및 경제적 여건으로 인해 정보수단에 대한 접근성의 배제로 인해 소외 현상을 의미한다. 이러한 정보격차는 정보화 초기부터 자주 논의되어 왔던 정보화의 대표적인 역기능으로 현재까지도 학술 및 정책 차원에서 논의되고 있다(Belanger & Carter, 2009; Katz & Aspden, 1997; Norris, 2001; Thierer, 2000; Warren, 2007). 정보격차 문제를 기술에 대한 접근성에 한정할 수는 없으나 대체로 접근성을 중심으로 해결하고 있는 상황이다. 이는 정보격차는 일반적으로 기술의 보유 현황 등 객관적인 통계에 기반하여 현황을 확인하고, 이에 따른 정책 처방도 직관적이고 그 결과를 파악하기 용이하다는 차이점이 있다.

반면 데이터 기반 사회가 야기하는 불평등은 현재까지 학술 및 정책적 논의가 초보적인 단계로 몇몇 연구에서 데이터 불평등에 대한 정의를 하고 있다. 하지만 연구자마다 접근 방식의 차이로 인하여 데이터 불평등에 대한 명확한 개념화는 다소 미진한 상황이다(Cinnamon, 2020; Kim, 2020; Lucas, et al., 2020; Lutz, 2019; Milan & Trere, 2020). 이러한 개념화의 문제는 기술에 대한 접근성을 중심으로 하는 기존 정보격차와 다르게 무형자산인 데이터가 초래하는 불평등이 다차원적으로 발생할 수 있으며, 불평등의 양상을 확인하고 특정하는데 한계가 있기 때문이다. 데이터의 수혜를 받거나 데이터를 활용하는 측은 데이터 자체가 편익을 제공하는 중요한 자원이라는 인식이 공고하기 때문에(Cinnamon, 2020; Seo, 2019), 데이터로 인한 역기능에 대해서는 상대적으로 관심이 저조한 편

이다. 그리고 데이터는 객관적이고 항상 옳다는 일종의 맹신도 데이터에 대한 비판적 접근을 약화시키고 불평등을 야기하는 요소가 될 수 있다(Kim, 2020; O'Neil, 2016).

그리고 정보격차가 특정한 의도를 가진 것이 아닌 최신 기술의 발전에 의한 불가피한 측면이 있다면, 데이터 기반 체계에 따른 불평등은 특정한 주체에 의해 의도적으로 광범위하게 발생 될 수 있다는 점에서 심각성을 인지할 필요가 있다. 데이터 기반 사회에서 데이터 확보가 어려운 사각지대의 존재는 기존의 격차를 고착화시킬 수 있기 때문이다. 데이터의 가치가 중요시되는 스마트 시티 구현에서도 이러한 데이터 중심의 도시 운영이 소외계층에 대한 격차를 고착화 및 증폭시킬 수 있다는 우려를 제기한다(New, et al., 2017; Reuter, 2020). 그리고 디지털 전환에 따른 공공부문의 데이터 수집 및 활용 역량의 증대로 인해(Seo, 2021), 최근 공공데이터의 가치가 증대되어 공공부문의 데이터 의존성을 더욱 가속화 시킬 것으로 전망된다. 오늘날 데이터 기반 사회가 필연적인 상황에서 이처럼 기존의 정보격차보다 더욱 광범위하고 복합적이며 문제점을 파악하기도 어렵다. 이에 따라 그에 따른 적절한 처방을 제시하는 것에 한계가 있다는 점에서 데이터 불평등 문제에 대한 이해와 해결을 고민할 필요가 있다.

이러한 문제인식을 기반으로 본 연구는 데이터로 인해 촉발될 수 있는 데이터 불평등에 대한 문제의 유형 및 사례 등을 탐색적으로 다루어 보고자 한다. 최근 디지털 기술의 확산에 따라 기존의 정보격차에 해당하는 디지털 기기의 보유에 따른 이슈는 지속적으로 다루어지고 있으나 데이터에 의한 격차 및 불평등에 대한 문제제기는 상대적으로 적은 편이다. 본 연구는 관련 문헌 및 국내·외 현황 자료를 중심으로 하여 데이터 기반 사회로 인해 야기 될 수 있는 격차 문제를 짚어보고, 디지털 전환이 포용적인 사회를 구현할 수 있도록 시사점을 도출하고자 한다.

II. 이론적 배경

1. 데이터 기반 사회의 출현

데이터(Data)는 특정한 행동에 대한 결정을 내리기 위해 활용되는 숫자, 글자, 그림, 기타 기록들의 구성을 의미한다. 데이터는 해석이 이루어질 때 의미가 부여되고, 정보(Information)로 인정된다(Richmond, 2006). 데이터 기반 사회의 출현은 ICT(Information Communication Technology)의 전방위적 확산에 따른 방대한 데이터의 도출과 이러한 데이터를 효과적으로 분석할 수 있는 인공지능, 빅데이터 등 데이터 기술의 발전에 기인한다. 데이터 기술은 개인, 그룹 등 사회 전반에 대한 추적, 예측, 통제 등을 가능케 하였고, 이에 따른 효율성으로 인하여 데이터를 중심으로 혹은 데이터에 의존하는 데이터 기반 사회가 등장하였다(Pentland, 2013; Wolff, et al., 2016). 디지털 기술을 활용한 데이터의 분석은 이슈에 대한 효과적인 통찰력을 제공하여, 생산성과 정확성을 증강시키게 되면서(Gartner, 2016; Seo, 2019), 데이터 기반 체계의 내재화는 현재 다양한 분야에서 필수적으로 고려되고 있다.

1) 민간부문의 데이터 활용

민간 비즈니스 분야는 고객과 시장을 파악하기 위한 목적으로 데이터 기반 체계 구축을 적극적으로 추진하고 있다. 비즈니스 모델 구축에 있어 데이터 활용이 필수적으로 언급되고 있는데, 일부 과도한 데이터 의존성에 대한 우려가 있지만, 기본적으로 데이터 기반 체계로 조직이 변모해야 한다는 것에는 이견이 없는 것으로 확인된다(Brownlow, et al., 2015; Bulger, et al., 2014; Hartmann, et al., 2014; Nuccio & Guerzoni, 2019; Sorescu, 2017). 비즈니스에서 데이터를 활용하는 것이 새로운 현상은 아니지만 빅데이터를 위시한 데이터 기반 비즈니스의 급속한 확산은 네 가지 측면에서 제시할 수 있다. 첫째는 규모(Volume)로 데이터의 양의 증대와 함께 데이터 저장기술이 저렴해지고, 데이

터 분석이 용이해졌다. 둘째는 다양성(Variety)으로 과거에는 스프레드시트와 같이 표준적인 내용을 제시하는 정형 데이터가 주류였으나, 최근에는 이미지, 동영상, 메일, SNS 등 사람들의 일상생활을 보여주는 비정형 데이터가 증대되고 있다. 셋째는 속도(Velocity)로 ICT(Information Communication Technology)의 발전으로 데이터의 수집·저장·분석이 보다 신속해지며, 이로 인해 데이터의 발생과 데이터의 활용 간의 시간적 격차가 감소하였다. 넷째는 정확성(Veracity)으로 신뢰성과 연관된 데이터 품질에 영향을 주는 요인들이 증대되면서 데이터를 엄격하게 관리할 필요가 나타났기 때문이다(Bulger, et al., 2014). 데이터 기반 비즈니스는 신생 스타트업과 기성 기업에게 수익 증대와 가치 창출을 가능케하고, 신규 비즈니스를 육성할 수 있는 혁신성을 제공할 수 있기 때문에 지속적으로 확대될 것으로 전망된다(Brownlow, et al., 2015; Hartmann, et al., 2014; Sorescu, 2017).

한편 데이터를 구축하기 용이한 온라인 기반의 디지털 플랫폼 비즈니스 기업이 데이터 기반 비즈니스를 선도하고 있다. 애플(Apple), 알파벳(Alphabet, Google의 모기업), 마이크로소프트(Microsoft), 아마존(Amazon), 메타(Meta, 전 Facebook) 등이 대표적인 글로벌 디지털 플랫폼 비즈니스 기업으로, 해당 기업들은 전 세계에 걸쳐있는 자사의 플랫폼을 기반으로 방대한 소비자들이 자발적 혹은 비자발적으로 제공하게 되는 데이터를 활용하여 고객의 수요를 파악하고, 이를 통해 수익을 극대화하고 있다(Bulger, et al., 2014; Nuccio & Guerzoni, 2019). 특히 디지털 플랫폼 비즈니스의 고객은 개인부터 기업, 지방정부, 중앙정부까지 다양하기 때문에(Sorescu, 2017), 이들 기업들이 가진 데이터의 가치는 플랫폼의 독점성에 따라 더욱 증대될 수 있다. 이에 플랫폼 기업들이 구축한 데이터 생태계에 대한 비판도 적지 않은 상황이다. 소수의 글로벌 플랫폼 기업은 이미 대중의 일상생활 및 업무와 밀접한 관련을 가지고 있어 배제가 어렵다. 또한 데이터 제공에 대한 규제정책이 있어도 기업들은 다양한 유인책으로 고객

이 자발적으로 데이터를 제공하도록 만들기 때문이다. 데이터 기반 비즈니스의 활성화가 기업의 수익 창출과 소비자의 다양한 수요충족이라는 긍정적 측면 이면에는, 글로벌 플랫폼 기업을 위시한 소비자의 플랫폼 종속화와 개인정보 침해와 같은 데이터 활용 범위의 문제 등 새로운 이슈를 야기하고 있다(Cinnamon, 2020; Kim, 2020; McCarthy, 2016; Schäfer, et al., 2023). 다만 디지털 환경이 보편화된 시점에서 규제 일변도의 정책을 제시하는데 한계가 있어 데이터 보호 및 활용의 적절한 타협점이 요구된다.

2) 공공부문의 데이터 활용

과거 전자정부 도입에 따른 행정전산화를 통해 정보의 접근성이나 신속성은 제고되었으나 당시에는 공공기관이 보유하고 있는 데이터의 가치에 대한 인식은 비교적 낮은 편이었다. 데이터는 특정한 목적에 한하여 축적되었으며, 무엇보다 데이터가 민간은 물론 공공기관 내에서도 공유가 용이하지 않아 활용이 제한적이었다. 공공기관의 데이터에 대한 관점이 변하게 된 것은 공공데이터 개방이 대두되면서 부터이다. 2009년 미국 오바마 행정부는 '투명하고 열린 정부'라는 국정기조를 위해 투명성, 시민참여, 협업 등의 세 가지 원칙을 제시하고 열린정부지침(Open Government Directive)을 발표한다. 특히 공공데이터와 관련하여 데이터 플랫폼인 Data.gov를 구축하고, 정부기관의 공공데이터 의무를 명시하였다는데 의의가 있다(Ginsberg, 2011). 공공데이터 개방의 효과는 공공부문에 대한 투명성과 효율성을 넘어, 복잡한 난제의 해결 및 혁신 등 비즈니스 창출과 같은 다양한 긍정적 효과 등이 제시되고 있다(Janssen, et al. 2012; Seo, 2017; Yannoukakou & Araka, 2014).

데이터에 따른 공공부문의 변화를 보여주는 것은 데이터 기반 행정(Data Driven Public Administration)이다. 데이터 기반 행정은 의견 기반 정책(Opinion-based Policy)에 대한 비판에서 시작된 증거 기반 정책(Evidence-based Policy) 개념에서 유래한다. 의견

기반 정책은 선택적인 증거의 활용 혹은 개인 및 그룹의 검증되지 않은 관점에 과도하게 의존하고, 이념, 편견, 어림짐작 등에 영향을 받는다(Davies, 2004). 반면 증거 기반 정책은 보다 합리적이고, 엄격하며, 체계적인 접근법으로, 활용 가능한 증거에 기반하여 많은 정보를 취득하고, 합리적 분석에 따라 정책결정이 이루어진다(Sutcliffe & Court, 2005). 데이터 기반 행정은 증거 기반 정책의 이념에 근거하며, 4차 산업혁명 및 디지털 전환으로 인한 데이터 기술의 진보와 공공데이터 개방과 활용에 대한 기반 구축 및 방향성으로 인해 등장한 공공부문의 업무방식 중 하나이다. 미국은 2019년 증거기반법(The Foundations for Evidence-Based Policymaking Act)을 제정하였다. 해당법은 연방정부 기관은 정책결정을 지원하기 위한 증거를 개발해야 하고, 이를 통해 정책이슈를 식별하기 위한 계획을 관리에 산실(Office of Management and Budget)과 의회에 제출해야 한다. 법률명에 데이터를 명시하지는 않았지만, 해당 법률은 2019년에 입법된 미국의 공공데이터 개방 기본법(OPEN Government Data Act)과 연계되어 있어, 데이터 기반 정책결정을 공식화한 것으로 이해할 수 있다.

2020년 제정된 국내 '데이터 기반 행정 활성화에 관한 법률(이하: 데이터 기반 행정법)'에서는 '데이터 기반 행정을 공공기관이 생성하거나 다른 공공기관 및 법인·단체 등으로부터 취득하여 관리하고 있는 데이터를 수집·저장·가공·분석·표현하는 등의 방법으로 정책 수립 및 의사결정에 활용함으로써 객관적이고 과학적으로 수행하는 행정'으로 정의하고 있다. 데이터 기반 행정은 데이터 기반 공공부문(Data-driven Public Sector)으로 표현되며, 데이터의 관리, 공유, 활용을 통해 공공정책 및 서비스에 대한 설계, 전달, 감시체계를 전환하는 것으로 정의된다. 이것은 데이터의 활용이 공공부문의 지능을 향상시키는 중요한 전략자산으로써 역할을 하기 때문이다. 결과적으로 장기적으로 지속가능하고, 가능한 포괄적이고 신뢰할 수 있는 정책과 서비스를 개발할 수 있는 역량을 증진시킨다(van Ooijen,

et al., 2019). 데이터 기반 행정은 피드백과 데이터 활용을 통해 지속적인 개선이 가능하고, 공공조직이 이용자의 요구에 대한 우선순위를 정하고 충족시키는 것을 가능케 한다(Ubaldi, et al., 2020). 국내의 경우 디지털플랫폼정부 추진에 따라 데이터에 대한 필요성이 더욱 증가하고 있는데, 기술적 측면에서 데이터 통합과 데이터 공유는 디지털플랫폼정부의 인프라 측면에서 핵심적이다(Seo, 2023b)

2. 데이터가 초래하는 불평등

디지털 전환으로 인해 사회 전반이 데이터를 중심으로 각종 의사결정과 서비스 등 제공 등이 이루어지고 있다. 데이터 기반 체계는 개인에 특화된 제품 및 서비스를 창출하여 다양한 개개인의 수요를 충족시킬 수 있다는 점에서 수요자 중심의 방향성에도 부합한다. 하지만 데이터가 부재한 상황에서 데이터 기반 체계는 맹점이 존재할 수밖에 없는데, 최근의 기술 패러다임 변화는 데이터에 대한 의존성이 높아지고 있다. 특히 공공부문도 데이터 기반 행정과 같은 데이터 중심의 정책 및 공공서비스 개발을 추진하고 있는 상황에서 데이터 부재 및 부족에 따른 문제제기는 거의 논의되고 있지 않다. 공공부문의 역할이 필요한 데이터 사각지대에 있는 소외계층은 이러한 데이터 기반 행정이 오히려 기존의 사회적·경제적 격차를 고착화 및 악화시킬 수 있다.

현재까지 정보격차와 같이 데이터의 격차 혹은 불평등으로 인한 학술적 차원의 공통된 용어는 확인되지 않고 있으며, 본 연구는 통일성을 기하기 위해 데이터 불평등으로 활용하고자 한다.

Britz(2004)는 정보빈곤(Information Poverty)에 대해 개인과 공동체가 정보에 대한 효율적인 접근이나 해석 및 적용하는데 필수적인 기술, 능력, 물적 수단 등을 가지지 못하는 상태로 정의한다. 정보빈곤은 필수 정보의 결핍, 정보인프라의 미비에 의해 사회적 국가적 수준으로 세분화 된다.

Sawyer(2008)는 과학기술 분야의 관점에서 데이

터 풍요(Data Wealth)와 구분되는 데이터 빈곤(Data Poverty)에 대해 지식축적 기반의 구축을 저해하고, 연구자들이 현상에 대한 거대 담론에 참여하는 역량을 제한한다고 하였다. 제한된 데이터로 인해 검증을 거치지 않은 호환되지 않은 데이터가 수집될 가능성이 높고, 이는 데이터 공유와 대규모 데이터셋 구축을 어렵게 할 수 있다.

McCarthy(2016)는 빅데이터로 인한 불평등 문제를, 단지 정보의 많고 적음이 아닌 다차원적인 관점에서 빅데이터 격차(Big Data Divide)로 제시하였다. 빅데이터 격차는 개인 혹은 그룹의 신원 및 실행 가능한 지식을 만드는 사람들과 그것의 대상이 되는 사람들 사이에서, 빅데이터 분석이 적용 가능한 그룹들 사이에서, 정보의 비밀유지가 있는 사람과 공개되는 사람들 사에서 발생하는 것으로 정의하였다.

Lutz(2019)는 정보격차의 단계를 접근, 활용, 성과 등으로 구분하고, 3단계 정보격차(Third-Level Digital Divide)인 성과측면에서 데이터 기반 알고리즘에 의해 고착화된 구조적 불평등을 강화하고, 이를 데이터 기반 차별(Data-based Discrimination)로 제시하였다.

Kim(2020)은 데이터 격차(Data Divide)를 크게 세 가지로 구분하였다. 데이터 생산 격차는 데이터에 접근하지 못해 발생하는 데이터 접근격차, 개인과 플랫폼 회사 간의 비대칭 문제로 인해 개인의 객체화 플랫폼 회사의 데이터 주체 문제인 데이터 통제격차가 있다. 데이터 표현 격차는 데이터에 대한 능동적 표현하는 사람과 표현하지 않는 사람간의 차이로 인해 발생하는 격차이다. 데이터 참여행위의 단계적 흐름과 성과의 격차는 데이터 활용 역량에 따른 격차를 의미한다.

Milan and Trere(2020)는 데이터 격차(Data Gap)를 데이터 생성과 데이터 품질에 관한 것으로, 데이터화가 취약한 사회에서 발생하는 이슈로 제시하였다. 데이터 격차는 데이터 빈곤(Data Poverty)을 야기하는데, 이것은 단지 데이터 소유의 많고 적음의 문제가 아닌 사회의 데이터화 이전에 본질적인 불평등에서 촉발된 것임을 지적하고 있다. 그리고 사회 전반의 데이터 인프라

로 인해 수치화 할 수 있는 대중에 대한 정책결정자의 의존성을 심화시켜 불평등을 악화시킬 수 있다.

Cinnamon(2020)은 데이터 불평등(Data Inequalities)을 접근, 표현, 통제 등으로 구분하여 설명하였다. 데이터 접근 불평등(Data Access Inequalities)은 데이터 중심 시대에 가장 가시적인 문제로, 사회·경제·지리적 요인으로 데이터 접근의 어려움이 촉발된다. 데이터 표현 불평등(Data Representation Inequalities)은 온라인 활동에 적극적으로 참여하거나 데이터를 생산하는 사람들이 그렇지 않은 사람에 비해 사회를 대표한다는 잘못된 인식으로 인해 발생하는 격차이다. 데이터 통제 불평등(Data Control Inequalities)은 데이터를 제공하는 측과 이러한 데이터를 수집·관리·활용하는 측간의 격차로, 오늘날 디지털 사회에서 의도하지 않게 개인 데이터를 기업 등에 제공하면서, 데이터를 제공한 개인은 자신의 데이터에 대한 권한을 잃는 문제가 발생한다.

Lucas, et al.(2020)은 기존의 정보격차 관련 개념으로는 데이터 불평등에 대한 설명에 한계가 있다고 전제하고, 데이터 빈곤(Data Poverty)라는 개념을 제시하였다. 데이터 빈곤은 개인, 가구, 공동체가 그들의 요구를 충족하기 위해 충분하고, 개인화되며 안전한 모바일 및 브로드밴드 데이터를 확보할 여건이 되지 못한 상태로 정의하였다. 데이터 빈곤은 가용성, 선택, 인프라, 개인정보보호 및 보안, 수량, 활용기술, 편의성 등의 차원에서 결핍 문제를 제시하였다.

Worldbank(2021)는 오늘날 데이터 기반 사회에서 저소득 국가의 경우 데이터 격차(Data Gap)로 인해 정책입안자의 올바른 정책의사결정과 시민들의 공공서비스 접근을 저해할 수 있음을 지적하고, 공공데이터 및 민간데이터의 격차로 인해 고소득 국가와 저소득 국가의 기존 경제적 사회적 격차가 더욱 증폭시킬 수 있음을 우려한다.

Giest and Samuels(2020)은 정부의 정책의사결정 과정에서의 데이터 격차(Data Gap)에 주목하고, 데이터 격차를 정책이 대규모 데이터 세트에 의존하는 상황

〈표 1〉 관련 문헌의 데이터 불평등 개념
 (Table 1) Concept of data inequality in literature

Researcher	Term	Contents
Britz(2004)	Information poverty	The difference in capability with regard to access, interpretation, and application of data
Sawyer(2008)	Data poverty	Contrary to data wealth, data poverty hinders the data application of researchers from the academic perspective
McCarthy (2016)	Big data divide	The gap occurs during the creation, management, and application of big data
Lutz (2019)	Data-Based Discrimination	Inequality problem based on the data dependence of the AI algorithm
Kim (2020)	Data Divide	Data divide is classified into data production divide, data representation divide, and data cascade and performance divide
Milan & Trere (2020)	Data Gap	Problem of creation and quality of data owing to low datafication of society
Cinnamon (2020)	Data Inequalities	The concept of data inequality is classified into access, representation, and control
Lucas, et al. (2020)	Data poverty	The status of being unable to acquire data to satisfy requests
Giest & Samuels (2020)	Data gap	The status of data on specific component or social group being intentionally or unintentionally omitted during the policy decision-making process
World Bank (2021)	Data gap	Data gap hinders policy decision making and public services provision in low- income countries
Ada Lovelace Institute (2021)	Data Divide	The divide between people who can access, choose, and control data based technologies and people who cannot that

에서 의도적 혹은 비의도적으로 특정한 요소 및 사회 그룹에 대한 데이터 누락된 상황으로 정의하였다. 데이터 격차를 세 가지 시나리오로 구분하였는데, ‘주요 데이터 격차(Primary Data Gap)’은 데이터가 누락을 인지하고도 정부가 누락된 데이터를 수집하지 않는 상황에서 발생한다. 숨겨진 데이터 격차는 정부가 의도하지 않게 대표성이 떨어지는 특정 그룹의 데이터 세트에 의존하면서 작은 규모의 잠재적 취약 그룹을 간과하면서 발생한다. ‘보조적 데이터 격차(Secondary Data Gap)’는 데이터가 있음에도 품질이 낮아 데이터 격차 해소에 적합하지 않는 경우에 발생한다.

Ada Lovelace Institute(2021)는 데이터 격차(Data

Divide)를 데이터 기반 기술에 접근·선택·통제할 수 있다고 여기는 사람과 그렇지 못한 사람간의 격차로 정의하였다. 데이터 격차는 데이터 기반기술을 대표하거나 접근하는 사람들에 의해 데이터 시스템의 설계, 개발, 형성되는 방식으로 명시됨으로써 정보격차와 상호 작용한다. 데이터 격차는 사회적·건강적 불평등을 고착화하고 악화시키게 된다.

연구자들의 데이터 불평등 관련 개념을 종합하면 〈표 1〉과 같이 정리할 수 있으며, 데이터 불평등에 대한 논의는 현재까지 주로 국외문헌에서 주로 확인되고 있다. 이는 국내의 경우 주로 PC에서 모바일로 이어지는 접근격차를 중심으로 한 정보격차에 초점을 두고 있기 때

문으로 보인다. 하지만 디지털 전환의 핵심은 결국 데이터의 활용에 있기 때문에, 기존의 정보격차 중심의 불평등 이슈를 조명하는 것은 근본적인 해결책이 되지 못한다. 한편 데이터 불평등에 대한 용어 및 정의 역시 구체적으로 통일된 접근은 확인되지 않는데, 이는 데이터와 관련된 불평등의 원인과 결과가 매우 광범위하기 때문으로 판단된다(McCarthy, 2016). 본 연구에서 보는 데이터 불평등은 데이터의 접근 및 전달(표현)의 장애 혹은 데이터의 부재 등으로 인해 데이터 기반 사회에서 사회적·경제적 소외가 발생하는 현상으로 정의하고자 한다. 데이터의 접근의 경우 기존의 정보격차 개념에서도 다루어지기 때문에 전달 및 부재의 측면을 중심으로 데이터 불평등을 확인하고자 한다.

III. 연구 설계

오늘날 디지털 전환으로 인한 데이터의 활용 가치는 더욱 증대되고 있으며, 일찍이 데이터 기반 비즈니스가 보편화된 민간부문 외에 공공부문도 데이터 기반 행정을 내세우는 등 데이터는 필수재로 부상하고 있다. 하지만 최신 기술 도입에 따른 정보격차 문제와 같이 데이터 기반 행정의 확대로 인한 새로운 차원의 불평등 이슈가 초래될 수 있다. 특히 민간부문과 비교하여 그 대상자의 범위가 넓고, 공공부문은 소외계층에 대한 지원에 핵심적인 역할을 하기 때문에 주의 깊은 그러나 현재까지 데이터에 따른 불평등·격차·소외 문제에 대해서는 정책적으로나 학술적으로 상대적으로 관심이 저조한 상황이다. 이는 기존 정보격차의 ICT(Information Communication Technology) 기기와 같이 데이터가 구체적인 형태를 가지지 않고, 그것의 격차로 인한 문제점을 확인하는 것도 한계가 있기 때문으로 보인다. 즉, 데이터 불평등의 현황을 판별할 수 있는 데이터가 부족하다는 것이다. 이는 현재까지도 국가간 혹은 국내의 정보격차를 측정요소로 ICT(Information Communication Technology) 인프라 및 기기 보유 등과 같은 접근격차 중심의 정량적 자료에 근거하는 것

과 궤를 같이한다(Seo & Myoeng, 2016). 이에 본 연구는 공공부문의 데이터 기반 행정이 보편화되는 상황에서 데이터의 수집, 관리, 활용 등의 단계를 중심으로 데이터 불평등 이슈를 시론적 차원에서 다루고자 한다.

첫째, 데이터 이용능력은 산재되어 있는 데이터를 수집하고 분석할 수 있는 능력을 의미한다. 현실 세계에 존재하는 다양한 현상을 데이터화(Datafication)하여 기계판독이 가능한 상태로 전환시키고, 이를 통해 현실 세계에 대한 이해를 디지털 기술을 활용하여 용이하게 할 수 있다. 데이터 이용능력은 기술적 차원에서 정보격차 개념과도 궤를 같이 하는데, 데이터 수집 및 분석을 위한 인프라는 곧 경제적 수준과 연결되기 때문이다. 해당 주제는 선행연구를 토대로 데이터 인프라의 편중 및 격차로 인해 발생하는 이슈를 중심으로 확인한다(Britz, 2004; Cinnamon, 2020; Giest & Samuels, 2020; Lucas, et al., 2020; Milan & Treere, 2020; Worldbank, 2021).

둘째, 데이터 관리는 데이터 활용을 위한 데이터 수집 및 활용 주체의 데이터 운영에 해당한다. 수집된 데이터의 축적 및 가공 등 데이터의 활용을 위한 전반적인 프로세스이다. 데이터 관리는 외부 이해관계자에게는 드러나지 않기 때문에 모니터링이 요구되지만, 이를 검증하는데 한계가 있다. 이로 인해 데이터 관리자측의 수집된 데이터에 대한 오남용 등의 우려가 제기될 수 있다. 해당 주제는 선행연구를 토대로 데이터의 관리자 측과 데이터 제공자 측의 권한, 정보 등의 불균형한 관계에 따른 격차를 중심으로 확인한다(Cinnamon, 2020; Kim, 2020; Lucas, et al., 2020; McCarthy, 2016).

셋째, 데이터 활용은 데이터를 통한 산출물(정책, 공공서비스, 의사결정, 제품 등) 등을 창출하는 단계이다. 데이터 중심 체계의 구축은 데이터를 활용하여 내부 업무 프로세스를 개선하고, 궁극적으로 수요자 중심으로 산출물을 개발에 목적이 있다. 데이터라는 객관적이고 체계적인 증거와 방대한 데이터를 분석할 수 있는 최신 디지털 기술의 활용에 따른 것이다. 하지만 앞서도 지적되었듯이 데이터의 결핍과 부재, 데이터의 편향으로 인해 데

〈표 2〉 데이터 불평등 분석틀
 〈Table 2〉 Framework of data inequality

Type	Contents	Unit of analysis by target	Reference
Inequality of data-using ability	Inequality based on data infrastructure that can collect and analyze data	- Inequality among citizens - Inequality among regions - Inequality among individuals and among countries	Britz (2004), Cinnamon (2020), Giest & Samuels (2020), Lucas et al. (2020), Milan & Trere (2020), World Bank (2021)
Inequality of data management	Inequality based on the asymmetric relationship between the data manager and the data provider	- Inequality between citizen and country - Inequality between citizen and corporation	Cinnamon (2020), Kim (2020), Lucas et al. (2020), McCarthy (2016)
Inequality of data application	Inequality based on output through data	- Inequality between citizens - Inequality between consumers	Milan & Trere (2020), Cinnamon (2020), Giest & Samuels (2020), Kim (2020), Lucas et al. (2020), Lutz (2019), World Bank (2021)

이터 기반 체계가 의도한 결과를 가져올 수 있는지 고려할 필요가 있다. 해당 주제는 선행연구를 토대로 데이터 기반 산출물을 통해 수혜를 보는 측과 그렇지 못한 측의 격차를 중심으로 확인한다(Kim, 2020; Milan & Trere, 2020; Cinnamon, 2020; Giest & Samuels, 2020; Lucas et al., 2020; Lutz, 2019; Worldbank, 2021).

본 연구는 각각 데이터 이용능력, 데이터 관리, 데이터 활용 등 세 가지 유형에 따른 데이터 불평등을 공공부문과 민간부문의 관련 사례를 통해 조명하고, 이에 따른 함의를 도출하고자 한다. 본 연구가 Cinnamon(2020), Kim(2020) 등 기존의 데이터 불평등의 유형화와 차별되는 부분은 선행연구들이 데이터 불평등 유형분류를 기존 선행연구의 개념을 종합하여 제시하고 있어 추상적이지만, 본 연구는 데이터 불평등에 대한 실제 사례를 중심으로 하여 구체적이고, 실무적 차원에서도 활용이 용이하다는 것이다. 이는 데이터 불평등 개념이 정보격차와 비교하여 관련 연구의 진행 및 인식의 미비로 인한 것으로 보인다. 이에 본 연구는 탐색적 연구로써 데이터 불평등의 정책이슈화를 위해 해당 개념의 구체화가 필요하다고 인식하였다.

데이터 불평등 사례의 경우 기존에 다루어진 정보격

차와 비교하여 데이터 불평등에 대한 정책 및 학술적 연구가 현재까지는 초보적인 단계로 기존의 사례별 분석 단위에 따른 분류에 다소 한계가 있다. 특히 데이터 이슈가 대체로 자본 및 권한을 지닌 정부 및 기업 등을 중심으로 발생되고 있음을 확인하였다. 이에 본 연구에서는 각 불평등 유형별 분석단위를 시민 간 불평등, 지역 간 불평등, 국가 간 불평등, 시민-국가 간 불평등, 소비자-기업 간 불평등, 시민 간 불평등, 소비자간 불평등 등으로 제시하고자 한다.

IV. 데이터 불평등의 유형 및 관련 사례

1. 데이터 이용능력

사회 전반의 요소를 데이터화하고 이를 수집 및 분석에 활용하기 위해서는 정보화 인프라 및 기기 등 기술의 내재화가 보편적으로 이루어져야 한다. 하지만 데이터 관련 기술의 적용은 정보격차 문제에서 제기되다시피 기존의 경제적·사회적 불평등의 연장선으로 이해될 수 있다. 경제적 사회적으로 소외된 개인 및 집단 등은 데이터 인프라 및 기기 등의 기술 구축에 상대적으로

배제되는 경향이 높기 때문에, 이들에 대한 데이터는 누락되거나, 데이터의 품질 저하로 이어지게 된다(Naude & Vinuesa, 2021). 한편 데이터 인프라 외에도 데이터 인프라 및 데이터를 보유하고도 데이터 이용 주체의 데이터 관련 지식 및 기술 등 역량에 따른 데이터 불평등이 발생할 수도 있다. 하지만 현재까지 이를 판별할 수 있는 공식적인 지표 등이 부재하다. 그리고 데이터 기반 사회에도 불구하고 기존의 고착화된 불평등으로 인해 데이터 인프라 격차는 더욱 극심해지고 있다. 이에 부득이 본 연구는 주로 데이터 이용능력의 불평등 사례를 다루면서 데이터 인프라 측면에 초점을 두었음을 밝힌다. 데이터 이용능력에 따른 불평등은 데이터 인프라에 대한 접근 주체를 기반으로 개인간 데이터 불평등, 지역간 데이터 불평등, 국가 간 데이터 불평등으로 분석단위를 구분하였다.

1) 시민간 데이터 불평등

스마트폰 등 모바일 기기는 통화, 문자 등 통신수단 외에도 정보취득, 여가, 재테크 등 다양한 목적으로 활용되고 있다. 특히 모바일 기기가 전국민에게 보편화되면서, 각종 온라인 서비스 이용 시 필요한 휴대폰 개인 인증의 핵심수단으로 활용되고 있다. 이처럼 모바일이 생활 필수재로 등장하면서, 더욱 빠르고 안정적인 모바일 환경 구현을 위한 이동통신기술도 급속히 발전하고 있다. 하지만 고령층 및 소외계층은 신규 모바일 기기

구매 및 신규 이동통신 서비스 가입률이 낮은 편인데, 이로 인해 모바일을 통한 각종 서비스 이용에 제약을 받는다는 점이다. 2019년 10월 제18호 태풍 미탁에 의한 강원도 삼척시에서 발생한 주택 사면 붕괴사고에 의한 인명피해는 재난문자를 수신하지 못하면서 발생한 사례로 지적된다. 삼척시가 재난문자를 발송했지만, 스마트폰이 아닌 2G폰을 주로 쓰는 고령층이 많았던 해당 지역의 주민들이 문자발송 사실을 인지하지 못하면서, 피해를 입었다는 것이다(News1, 2019). 2G폰 외에도 3G폰도 기술적 문제로 재난문자를 받지 못하는데, 재난 문자 사각지대 해소를 위해 2017년 행정안전부는 안전디딤돌 앱을 배포하였으나, 2G와 3G의 경우 스마트폰이 아니기 때문에 앱을 사용할 수 없다는 문제점이 있었다(Money Today, 2023). 과학기술정보통신부의 이동통신기술별 회선수 현황을 보면, 2023년 12월 기준으로 2G의 경우 급격히 감소하여 가입자 수가 없으나 3G는 여전히 210만 회선에 달하는 서비스 이용자가 있는 것으로 확인된다(Ministry of Science and ICT, 2023). 그리고 재난 문자 외에도 앱을 활용한 각종 서비스의 사각지대에 놓여있는데, COVID-19 당시를 예로 들면 스마트폰이 없는 경우 마스크 재고알리미 앱, 백신 패스 등의 이용 제약으로, 일상생활에 적지 않은 피해를 볼 수 있다. 그리고 공공과 민간에서는 스마트폰을 이용자와의 주요 소통창구로 활용하고 있는데, 정책·제품·서비스 개발을 위한 이용자의 데이터를 취득하기 보

〈표 3〉 국내 이동통신기술별 연간 회선 수

〈Table 3〉 Number of lines per year by mobile telecommunication technology (Korea)

	2018.12	2019.12	2020.12	2021.12	2022.12	2023.12
2G	1,672,741	1,020,510	502,585	171,930	51,115	0
3G	9,549,356	7,515,903	5,604,557	3,479,622	2,664,193	2,134,958
4G	55,133,681	55,687,974	52,555,161	48,288,764	46,217,456	48,948,694
5G	0	4,668,154	11,851,373	20,915,176	28,059,343	32,808,121

source: Ministry of Science and ICT (2024)

다 용이하기 때문이다. 하지만 스마트폰이 없는 고령층 및 소외계층은 그들의 의견을 표출할 기회를 얻지 못하여, 이들의 수요를 반영하는데 한계가 발생할 수 있다.

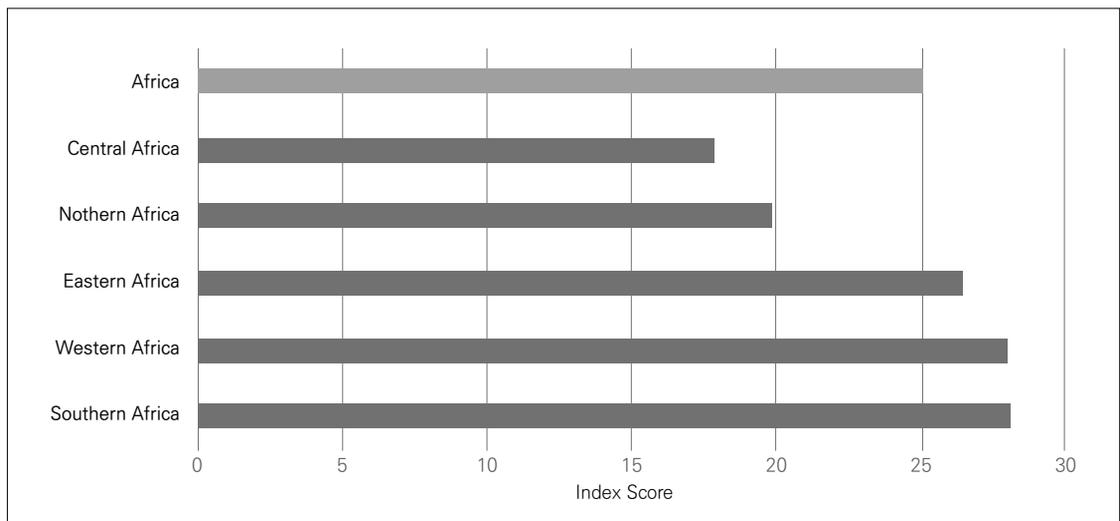
2) 지역 간 데이터 불평등

2022년 10월 국회 과학기술정보방송통신위원회 국정감사에서 이정문 의원이 과학기술정보통신부의 자료를 토대로 국내 17개 광역시·도별 공공와이파이 100km² 면적대비 공공와이파이 구축현황을 보면, 가장 많은 공공와이파이 구축지역은 서울(948개소), 대전(397개소), 광주(378개소), 인천(300개소)등이며, 가장 낮은 공공와이파이 구축지역은 강원(14개소), 충북(20개소), 전북(21개소)등으로 확인되었다. 이러한 지역 간 편중의 원인은 각 지자체의 재정여건에 따라 공공와이파이의 구축 수요에 차이가 나타나기 때문이다. 공공와이파이는 통신비 부담을 완화할 수 있다는 측면에서 공공와이파이 구축 미비 지역을 중심으로 통신소외 계층의 증대 우려가 제기되었다(The Seoul Shinmun Daily, 2022). 한편 데이터 수집에 있어 공공와이파이

의 중요성은 와이파이 핑거프린트(Wi-Fi Fingerprint)를 통해 위치정보를 수집할 수 있다는 것으로, 이는 정책 및 공공서비스 개발에 유용하게 활용될 수 있다. 따라서 공공와이파이 구축의 미비는 정책 및 서비스에 활용할 데이터의 부족을 야기할 수 있다. 그리고 단순히 인프라 구축의 유무를 넘어서 시민들이 지역 내 삶의 질과 연관될 수 있기 때문에 지역 간 불균형을 초래할 수 있다.

3) 국가간 데이터 불평등

COVID-19 당시 핸드폰의 통신기록(Call Detail Record) 데이터는 인구밀집 수준을 판단하여, 지역 봉쇄 및 감염경로 추적 등 감염병에 대한 효과적인 대응을 가능케하였다. 하지만 통신기록 데이터 활용의 경우에도 형평성에 대한 문제가 존재한다. 예를 들어 고소득 국가인 이스라엘은 국민의 스마트폰 이용률이 92%이나, 저소득 국가인 감비아는 75%에 불과하다. 스마트폰 보유 인구의 부재로 인해 방역에 활용할 수 있는 통신기록 데이터를 얻을 수 없고, 이를 활용한 방



source: PARIS 21 & MIF(2021: 10).

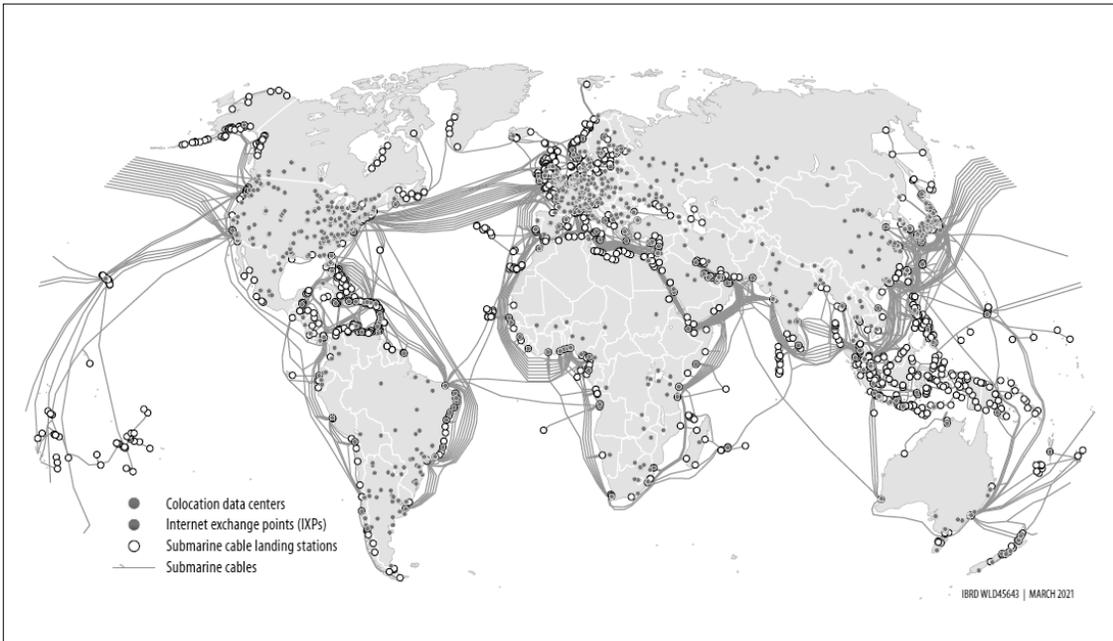
〈그림 1〉 아프리카 국가들의 정책결정의 낮은 통계정보 참조 수준

〈Fig. 1〉 Level of reference of statistical information in the policy decision making of African countries

역대응도 한계를 가질 수밖에 없게 된다(Worldbank, 2021). Naude & Vinuesa(2021)는 데이터 관점에서 COVID-19 발생에 따른 전세계적 교훈을 다루면서, COVID-19는 기존의 데이터 및 정보통신기술에 대한 접근격차만이 아닌 데이터 기반 정책결정 차원에서 데이터를 분석할 수 있는 역량 문제도 강조하였다. 저개발국가의 경우 이용가능한 데이터의 부재, 분석역량 외에도 저품질의 데이터로 인한 이른바 통계적 비극(Statistical Tragedy)이 팬데믹을 통해 확인되었음을 지적하였다.

국가의 통계역량개발을 촉진하기 위해 UN(United Nation), EU 집행위원회(European Commission), OECD(Organization for Economic Cooperation and Development), IMF(International Monetary Fund), Worldbank 등이 공동 설립한 단체인 PARIS21(Partnership in Statistics for

Development in the 21st Century)과 아프리카 자선 재단인 MIF(Mo Ibrahim Foundation)의 2021년도 보고서에 따르면, COVID-19는 증거 기반 정책결정을 위한 정비된 데이터 시스템의 중요성을 보여주지만, 아프리카 등 개발도상국은 데이터 축적의 미비로 인해 한계가 있음을 강조한다. 보고서는 국가의 데이터 구축을 위한 통계 역량(Statistical Capacity)은 해당 국가의 거버넌스 수준과 비례함을 보이며, 고품질 통계를 제공하는 국가는 더 나은 공공재와 공공서비스를 시민들에게 제공함을 언급하였다. 하지만 아프리카 국가는 데이터 인프라의 미비로 통계 자료의 품질이 떨어지고 누락되는 경우가 많았다. PARIS21의 2019년 조사에 따르면 통계정보가 정책결정 및 정책 모니터링에 사용되는지 여부를 100점 만점으로 분석한 결과, <그림 1>과 같이 아프리카 국가는 평균 25점으로 낮은 편에 속하였다. 데이터의 미비로 인해 COVID-19는 아프리카의 근



source: World Bank (2021).

〈그림 2〉 전세계 데이터 인프라 구축현황
 〈Fig. 2〉 Status of establishment of world data infrastructure

본적인 데이터 격차를 악화시켰는데, 의료시설 및 의료 정보에 대한 열악한 데이터 보장 범위로 인해 응급상황에 필요한 통계 자료 생산과 감염병에 대한 대응과 회복을 어렵게 하였다(PARIS 21 & MIF, 2021).

개인의 기기 보유 외에도 국가 단위의 데이터 인프라 수준으로 인한 국가 간 데이터 불평등 문제도 발생할 수 있다. 저소득·중소득 국가는 자체적으로 데이터를 생산하여 교환하고, 저장하며, 처리하는 국내 인프라가 부족하다. <그림 2>는 2021년 기준 전세계 데이터 인프라 관련 코로케이션 센터, 인터넷 익스체인지 포인트, 해저케이블 스테이션, 해저케이블을 구축현황에 대한 도식화를 나타낸 것으로, 데이터 인프라가 특정 대륙 및 특정 국가에 편중되어 있음을 확인할 수 있다. 데이터 인프라가 부족한 경우 해외 데이터 시설을 통해 데이터 전송할 수 있으나, 이 경우 낮은 속도 및 높은 비용 등 상당한 불이익을 감수해야 한다(Worldbank, 2021).

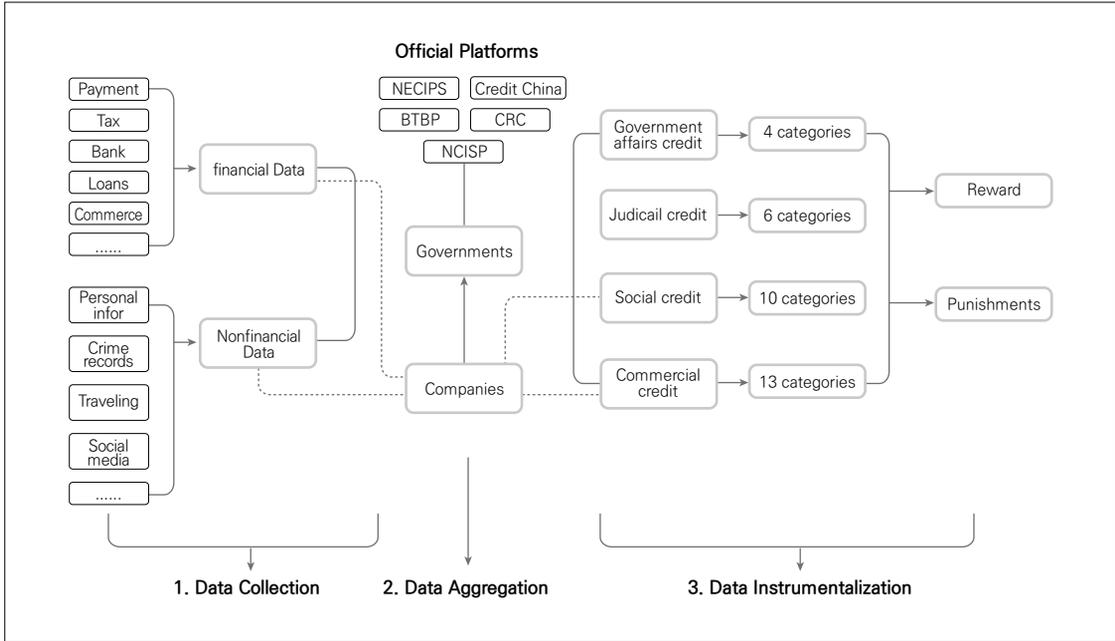
2. 데이터 관리

데이터 기반 사회에서는 수집된 데이터를 관리하는 주체의 영향력이 극대화되고 있다. 공공부문은 이미 국민에 대한 독점적인 정보권력을 가지고 있는데, 지능정보기술의 발전을 통해 행태데이터 등 보다 다양한 유형의 데이터를 보유할 수 있는 역량이 확대되었다(Hartzog & Selinger, 2015; Seo, 2021). 민간부문의 경우 민간 플랫폼 기업의 서비스를 통해 소비자들의 개인 데이터를 의도적 혹은 비의도적으로 제공하는 경우가 증대되었고, 이들이 남긴 개인 데이터를 가치를 창출하는 글로벌 플랫폼 기업의 영향력이 증대되고 있다(Kim, 2020; McCarthy, 2016). 빅데이터가 활용되는 상황에서 온라인 및 디지털 기기와의 접촉은 개인의 행동, 선호, 소망 등 다양한 데이터를 데이터 수집 주체에게 의도하지 않게 제공할 수 있다(Cinnamon, 2020). 하지만 데이터를 통제하는 위치에 있는 주체에 대해 일반 개인은 자신의 데이터에 대한 권리를 주장

하기 어렵기 때문에, 이는 비대칭적이고 불균형한 관계로 이어진다(Hartzog & Selinger, 2015; Kim, 2020; McCarthy, 2016; Schäfer, et al., 2023). 한편 데이터 관리자 간에도, 예들 들어 중소기업과 대기업, 데이터 활용이 용이한 국가와 그렇지 못한 국가에 데이터 수집, 활용에 대한 격차가 발생할 수도 있는데 본 연구는 데이터 제공자와 데이터 관리자 간의 갑을관계가 사회적으로도 문제가 될 수 있다고 판단하여 해당 사례에 집중하였다. 데이터 관리에 따른 불평등은 데이터 제공자와 데이터 관리자의 관계를 중심으로 시민-국가 간 데이터 불평등, 소비자-기업 간 데이터 불평등으로 분석 단위를 구분하였다.

1) 시민-국가 간 데이터 불평등

사회신용시스템(Social Credit System)은 중국 정부가 운영하는 빅데이터 기반 국가통합 데이터 플랫폼이다. 2007년 중국 장관연석회의에서 사회신용시스템 구축에 대한 논의가 시작되어, 2011년 정부분야, 소비분야, 사회활동, 사법분야 등 네 가지 분야를 근간으로 하는 초안이 마련된다. 2014년 중국 국무원은 사회신용시스템 구축 계획(2014~2020)을 발표하며 공식화된다. 사회신용시스템은 평가와 감시를 목적으로 공공과 민간 분야의 금융 데이터(은행내역, 세금, 대출, 거래 등) 및 비금융 데이터(취업, 교육, 범죄기록, 소셜미디어 이용 등) 등 방대한 데이터가 수집된다. 시스템에 통합된 데이터는 네가지 유형에 따라 분류되어 빅데이터를 통해 신용점수를 도출하는데 활용된다. 신용점수는 평가대상(개인, 기업, 단체, 공공기관 등)에 대한 보상 혹은 처벌에 대한 운영방식의 근간으로 활용된다. 평가대상이 신용점수가 좋고, 공산당의 규정을 잘 준수하면 보상을 받는 레드리스트에 등재되고, 반대로 신용점수가 낮으면 각종 사회·정치·경제적으로 제약을 받는 블랙리스트에 등재된다(Liang, et al., 2018). 사회신용시스템의 논란이 되는 부분은 방대한 데이터 수집에 따른 프라이버시 논란 외에도 개인의 선호 및 성향 또한 국가의 통제하에 놓일 수 있다는 점이다. 한 예시로 게임을 구



source: Liang, et al.(2018: 427)

〈그림 3〉 중국 사회신용시스템의 운영 개요
 (Fig. 3) Operational overview of China's social credit system

매하는 사람보다 기저귀를 구매하는 사람이 신용점수를 더 높게 받게 된다(Kim, 2017). 사회신용시스템의 구축의 옹호 의견으로 사회 투명성 및 신뢰의 증진이 제시되나, 중국정부의 사회 전반에 대한 통제력을 강화를 위한 역할이 적지 않다(Lee, 2019).

COVID-19 팬데믹 당시 기술의 활용은 방역대응에 긍정적 역할을 했으나(Seo & Myeong, 2021), 공익을 위한 국가의 통제 및 감시와 개인 자유의 대립은 통제지향적인 국가는 물론 자유주의 성향의 국가에서도 논쟁을 야기할바 있다. 방역 대응을 위한 기술 중 전세계적으로 가장 보편적인 것이 스마트폰 위치정보 및 앱을 활용한 감염자 확인 및 감염경로의 추적이다. 하지만 방역 대응을 목적으로 한 위치정보의 활용은 감염병과 무관한 개인 민감정보에 대한 노출 및 보안 우려가 제기되었다. 정부는 감염병 대응을 목적으로 시민에 대한 광범위한 데이터를 합법적으로 취득할 수 있

게 되면서(Smidt & Jokonya, 2021), 전례 없는 공공기관의 데이터 권력이 제고된 것이다. 일례로 국내에서는 COVID-19 초창기 감염자의 이동 동선이 홈페이지를 통해 공개되면서 과도한 개인 사생활 침해 논란이 야기되었다. 위치 데이터만으로도 특정 개인의 생활패턴 및 관심사에 대한 상당한 정보를 제공해주기 때문이다(Seo, 2021). 그리고 COVID-19에 대한 대응으로 웨어러블 기기 착용이 증대되면서, 이를 활용한 인권 및 자유에 대한 억압 및 통제 우려가 제기되기도 하였다. 공공보건을 위한 기기의 도입이 새로운 Big Brother를 잉태하였다는 것이다(Raposo, 2021). 한편 COVID-19 당시의 첨단기술을 활용한 방역대응은 해당 기술을 공공부문에 제공하는 관련 기업의 데이터 수집 및 분석 역량도 제고시켰다(Qi, 2020). 이 과정에서 시민의 민감 데이터가 민간기업에 의해 무분별하게 유출될 우려가 발생할 수 있다.

2) 소비자-기업 간 불평등

구글(Google)의 모기업 알파벳(Alphabet)의 자회사인 사이드워크랩스(Sidewalk Labs)는 2018년 캐나다 정부, 온타리오 주정부, 토론토시가 설립한 Waterfront Toronto와 토론토 퀘이사이드 지역에 스마트시티 구축 프로젝트는 Sidewalk Toronto를 위한 파트너십을 맺는다. 하지만 2019년 캐나다 인권단체인 CCLA(Canadian Civil Liberties Association)가 사이드워크랩스(Sidewalk Labs)의 무분별한 데이터 수집 및 개인 사생활 침해, 정보유출을 근거로 사이드워크랩스(Sidewalk Labs)와의 사업무효 소송을 주정부를 상대로 제기하게 된다. 이에 사이드워크랩스(Sidewalk Labs)는 개인정보이슈에 대응하기 위한 세 가지 방안으로 독립기구인 Urban Data Trust 설립, Responsible Data Use 가이드라인 수립, 도시 데이터 사용 및 수집을 위한 명확한 프로세스 설립 등을 제안하고, 2019년 10월 조건부로 승인되었으나 2020년 5월 최종적으로 사업추진이 철회되었다(Austin & Lie, 2021; Lee, 2020). 사이드워크랩스(Sidewalk Labs)의 토론토 스마트시티 프로젝트 실패는 전 세계적인 COVID-19로 인한 대외환경의 불투명성도 있으나, 플랫폼 기업의 데이터 활용에 대한 신뢰 문제가 주요 요인 중 하나로 제기된다. Sidewalk Labs는 토론토 외에도 미국 오리건주 포틀랜드시와 2019년 스마트시티 프로젝트를 추진하였다. 자회사인 레플리카(Replica)를 통해 사람들의 이동방식을 매핑하여, 축적된 데이터를 모빌리티 서비스 개발에 활용하고자 했다. 하지만 포틀랜드시와 시민들이 Replica의 데이터 수집 및 활용의 투명성에 문제제기를 하면서 2021년 중단되었다.

그리고 민간 플랫폼 비즈니스 확대에 인한 데이터 불평등 문제도 이슈로 부상하고 있다. 오늘날 많은 소비자들이 제품 및 서비스 이용에 웹사이트 및 앱 등의 디지털 플랫폼에 의존하고 있다. 플랫폼 기업들은 소비자의 고객 경험(Customer Experience)을 증진한다는 목적으로 가급적 소비자의 다양한 데이터를 보다 많이 확보하고자 한다. 맥킨지 앤 컴퍼니(McKinsey &

Company)의 미국 소비자 1,000명을 대상으로 한 자체 조사에 따르면 하지만 소비자의 입장에서는 본인이 이용하는 제품 및 서비스에 대한 정보만을 요구하는 기업에 대한 신뢰가 높은 것으로 나타났다(Anant, et al., 2020). 하지만 플랫폼 비즈니스에서 플랫폼 관리자와 플랫폼 이용자 간의 비대칭성으로 인해 플랫폼 이용자가 플랫폼 기업의 데이터 이용 및 관리 현황에 대해 면밀히 파악하기 어렵다. 이에 주요 플랫폼 기업들이 이용자의 데이터 무단 사용에 대한 위반 사례가 적지 않다. EU의 개인정보보호법에 해당하는 GDPR(General Data Protection Regulation)을 위반한 주요 플랫폼 기업의 사례를 보면 아마존(Amazon)은 2021년 이용자의 쿠키를 통해 개인 데이터를 수집 및 공유하면서 7억 4600만 유로의 벌금을 부과 받았고, Facebook을 운영하는 메타(Meta)는 2020년 Facebook 이용자의 개인 데이터가 온라인 해킹 포럼에 공개되어 적절한 데이터 관리 조치를 취하지 않았다는 이유로 2억 6500만 유로의 벌금을 부과 받았다(Tessian, 2022). 하지만 소수의 빅테크 기업에 의해 디지털 플랫폼 생태계가 독점적으로 운영되면서, 이들 기업의 데이터 독점에 대한 우려는 전세계적으로도 높은 상황이다. 개인 데이터 이용에 대한 규제에도 불구하고, 플랫폼 독점 상황에 대한 견제 기능이 부재하여 많은 소비자들이 소수의 플랫폼 기업에 자발적 혹은 비자발적으로 데이터를 제공하기 때문이다(McIntosh, 2018). 민간 사이버보안 업체인 클라리오(Clario)는 전세계적으로 많이 사용되는 일부 앱에 대한 분석 결과를 통해 이용자로부터 가장 많은 데이터를 취득하는 앱을 분석하였고, 상당수 유명 플랫폼 기업의 앱이 높은 순위에 있었다. 이들 앱은 정보접근에 대한 한번의 '동의' 만으로 이미지 및 음성 등의 데이터를 취득하고 이용자의 사생활에 쉽게 접근한다는 문제점이 제기되었다(Slynchuk, 2022).

3. 데이터 활용

데이터 기반 체계에서 데이터를 활용하는 것은 더 나

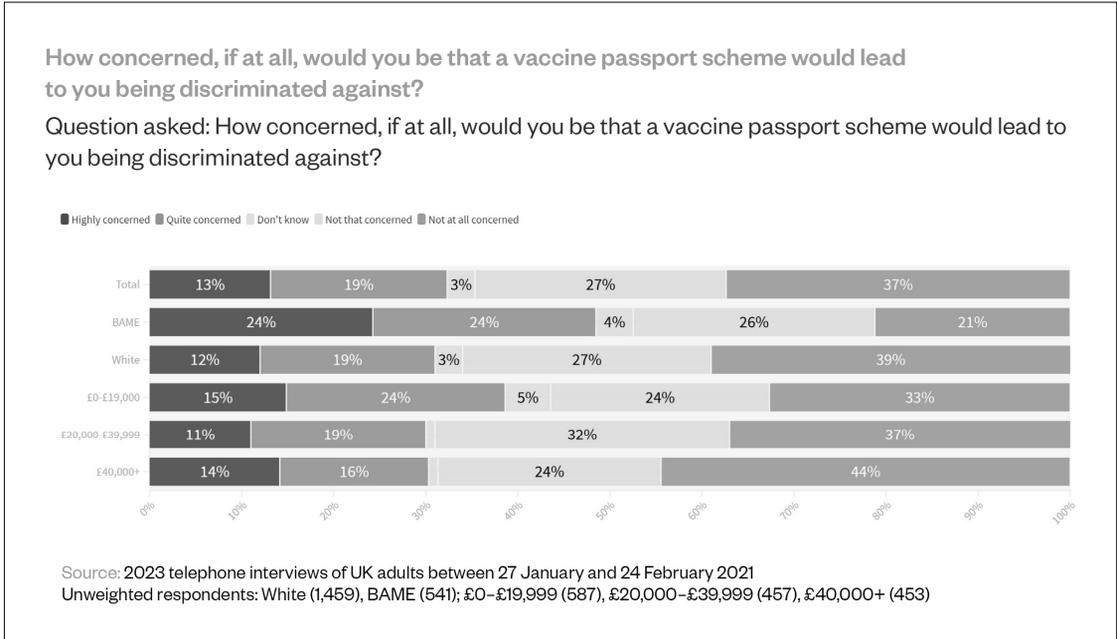
은 의사결정과 제품 및 서비스를 도출하기 위한 목적이 있다. 이는 데이터에 근거하는 것이 합리적이고, 과학적이기 때문에 편향성을 배제할 수 있고, 이를 통해 실제 수요자의 필요를 정확하게 파악할 수 있다는 인식에 근거한다(Seo, 2019). 하지만 최근의 데이터 중심 체계가 오히려 데이터에 대한 과도한 의존성을 촉진하여, 모든 현상을 데이터에 근거하여 해석하려는 데이터 만능주의로 오도될 우려가 있다. 이는 데이터는 현실세계를 그대로 대변하지 못하며, 데이터에 대한 의존으로 오히려 기존의 격차를 고착화시키거나 새로운 경제적·사회적 격차를 야기할 수 있음을 경계해야 한다는 것이다(Giest & Samuels, 2020). 구체적으로 데이터 활용에서의 격차는 특정 계층에 대한 데이터의 부족이나 편향 등으로 인해 데이터 기반 체계의 혜택을 받지 못하거나 차별을 당하는 경우를 의미한다. 데이터 활용에 따른 불평등은 데이터에 따른 수혜자를 중심으로 시민간 데이터 불평등, 소비자간 데이터 불평등으로 분석단위를 구분할 수 있다.

1) 시민간 데이터 불평등

인공지능 안면인식 기술은 효용성에도 불구하고, 인공지능 알고리즘의 편향성으로 인한 우려를 보여주는 대표적인 사례에 해당된다. 이러한 편향성의 원인은 상당 부분 데이터의 부족에 기인한 것으로(Mehr, 2017), 인종 및 성별 등 비주류 혹은 사회적 약자 등에 대한 차별로 이어지는 문제점이 있다. 2015년 구글(Google)의 포토앱은 흑인 사진에 ‘고릴라’로 라벨을 붙인바 있고, 2020년 트위터(Twitter, 현 X)의 자동 이미지 자르기 및 이미지 미리보기에서 해당 인공지능이 흑인보다는 백인 이미지를 노출하려는 경향이 높은 것으로 나타났다. 2021년 메타(Meta)의 페이스북은 동영상 자동추천 기능을 통해 흑인이 등장하는 영상에 ‘영장류에 관한 영상을 계속 시청하시겠습니까(Keep seeing videos about Primates)’라는 메시지를 게시하였다(BBC, 2021). 2020년 BBC 조사팀은 영국 내무부(Home Office)가 신속하고 편리한 여권 발급 절차를 위해 도

입한 온라인 여권 발급시스템의 증명사진 자동인식 기능의 인종적 편향성을 지적하였다. 어두운 피부색의 사진에 대해 정확성이 떨어지고, 해당 사진을 낮은 품질로 인식할 가능성이 높다는 것으로, BBC 조사팀은 전세계 1,000명 이상의 정치인 사진을 비교 분석에 활용하였다. 그 결과 밝은 피부색 여성의 14%가 저품질 사진으로 인식되는 반면, 어두운 피부색 여성의 22%가 저품질 사진으로 인식되었다. 밝은 피부색 남성의 9%가 저품질 사진으로 인식되는 반면, 어두운 피부색 남성의 15%가 저품질 사진으로 인식되었다. 일례로 여권사진 규정에 따른 흑인 여성의 사진을 자동인식 기능이 입을 벌리고 찍었다는 오류 판정하였다(BBC, 2020). 공공부문과 민간부문에서 공통적으로 확인된 안면인식 기술의 오류 문제는 특히 소수인종일수록 피해를 볼 수 있음을 보여주고 있다. 이는 알고리즘 학습에 활용된 데이터가 주로 백인 등 주류 계층의 이미지를 대상으로 학습하면서 발생한 편향성의 문제에 해당하며, 관련 업계의 편향성이 차별과 격차를 보여주고 있다.

인간과 사회를 중심의 인공지능 및 데이터를 목적으로 하는 영국의 민간연구소인 에이다러브레이스연구소(Ada Lovelace Institute)는 영국 건강관리 자선단체 Health Foundation과 협업하여 2021년 COVID-19 데이터 기반 기술 관련 인식을 중심으로 영국 성인 2,023명을 대상으로 한 전화 인터뷰를 진행하였다. 문항 중 백신패스(Vaccine Passport)의 도입이 자신에게 차별을 초래할 수 있다는 인식에 대해 흑인, 아시아인, 기타 소수인종 계층과 연소득 20,000 파운드 미만의 계층에서 차별에 대한 인식이 각각 백인 및 20,000 파운드 소득 이상의 계층보다 높게 나타나는 것으로 확인되었다. 보고서에서는 기존의 경제·사회 불평등이 데이터 기반 기술 활용에서도 전이되고 있음을 언급하였다(Ada Lovelace Institute, 2021). 다만 구체적인 원인에 대해서는 밝히지 않고 있는데, 데이터 불평등 관점에서는 백인 외 인종 및 저소득층의 경우 상대적으로 데이터 미비로 인해 COVID-19에 대한 정부의 방역대응 사각지대에 있을 가능성이 높고, 따라서 백신패스를 보유



source: Ada Lovelace Institute (2021)

〈그림 4〉 영국 시민대상 백신패스 도입의 차별 초래에 대한 인식

〈Fig. 4〉 British citizens' perception on concerns regarding the introduction of vaccine passport

하지 못할 가능성이 높아 기존의 고착화된 불평등을 더욱 증폭 될 수 있을 추론할 수 있다.

2) 소비자간 데이터 불평등

디지털 전환시대에서 데이터의 양이 기하급수적으로 증가하면서, 데이터 분석이 공공과 민간분야 모두에서 확대되고 있다. 하지만 많은 사례에서 데이터 분석이 실패하고 있는데, 핵심 원인은 투입 요소인 올바른 데이터의 부재가 지적된다(Hotz, 2022). 아마존(Amazon)은 인재 채용의 기계화를 위해 2014년부터 구직자 이력서를 검토하는 별도의 프로그램을 구축하였다. 해당 시스템은 인공지능에 기반하여 구직자에게 별을 기준으로 점수를 부여하여 선발하도록 되었는데, 2015년 해당 시스템의 성편향성이 확인된 것이다. 채용 프로그램은 남성을 여성보다 더욱 선호하는 것으로 나타났고, 여성이 포함된 이력서는 불이익을 주었다. 또한 성별이 포함

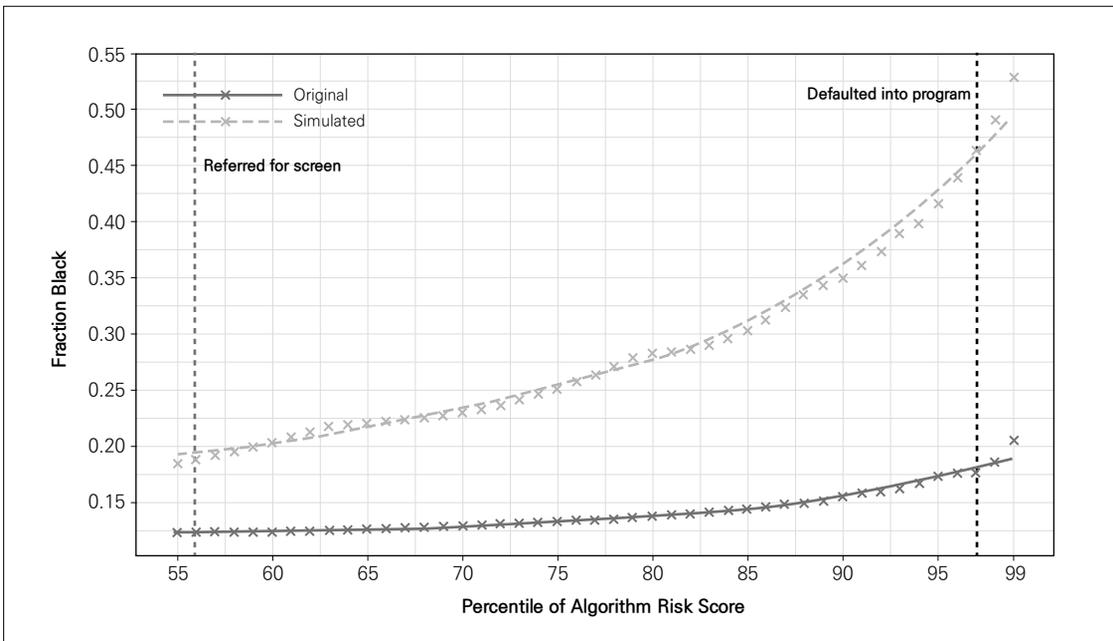
되지 않아도 남성이 주로 쓰는 언어에 대해 가점이 높은 것으로 확인되었다. 이러한 편향은 과거 이력서 패턴을 조사하면서, 아마존(Amazon)의 지원자 상당수가 남성이었기 때문으로 확인되었다(Reuters, 2018). 마이크로소프트(Microsoft)는 2016년 3월 23일 18-24세 미국 여성의 특징을 지닌 인공지능 챗봇 Tay를 실험적으로 트위터(Twitter, 현 X)를 통해 실험적으로 출시하였고, 이후 93,000개 이상의 트윗을 생산한 Tay를 하루 만에 중단시켰다. 당초 농담·시·낭송·괴담 공유·음악 가사 게시 등 엔터테인먼트 목적으로 만들어진 Tay는 특정 커뮤니티 이용자들이 악의적으로 언어를 학습하도록 유도하면서 공격적이고, 부정적인 언어를 쓰도록 변모한 것이다. Tay는 인종·정치·성에 대한 편향적 트윗을 게시하면서 인공지능에 대한 논쟁을 촉발하였다. Tay의 사례는 실질적으로 인공지능에 대한 문제이기 보다는 인공지능이 활용하는 데이터의 품질과 포

팔성이 중요함을 보여준다(Neff & Nagy, 2016).

유럽의회(European Parliament)의 여성의 권리와 성평등 위원회(European Parliament Committee on Women's Rights and Gender Equality)의 권고에 따라 유럽의회의 시민권 및 헌법 정책부(Policy Department for Citizens' Rights and Constitutional Affairs)에서 수행한 2023년도 연구 보고서에서는 AI, 로봇 등 디지털 기술의 확대에서 성별 구분 데이터(Gender Disaggregated Data)의 부족으로 여성, 트랜스젠더, 기타 취약계층이 기술 및 사회 발전에 소외되어 불이익을 받을 수 있음을 우려하고 있다. 성별 구분 데이터 부재로 인한 교통 및 공공시설, 보건, 재정 및 세금, 소비재 등 다양한 분야에서 발생할 수 있는 불평등 문제를 언급하고 있다. 소비재 측면에서는 다수의 상품이 남성을 기반으로 디자인되어 있

어 편향성이 지적되는데, 이는 여성 및 트랜스젠더 등에 대한 데이터의 부족, 여성 상품 디자이너의 낮은 대표성, 그리고 데이터를 활용하는 ICT(Information Communication Technology) 및 공학 관련분야의 여성 진출의 저조함 등이 요인으로 지적되었다. 해당 보고서는 성별 구분데이터에 따른 제품 및 정책설계의 필요성을 강조하며, 성별 민감 지표(Gender Sensitive Indicators)를 고려할 것을 제시하였다(Davaki, 2023).

의료분야의 데이터 누락 및 데이터 기반 의존으로 인한 격차 문제도 대두되고 있으며, 이는 기존의 경제·사회적 격차가 생명과 직결되는 건강 격차를 야기할 수 있음을 보여준다. 의료분야의 환자 진단 의사결정에서도 데이터에 따른 편향이 발생할 수 있는데, Obermeyer, et al.(2019)는 미국의 병원 및 보험회사에서 활용되



source: Obermeyer, et al. (2019: 3).

〈그림 5〉 의료 알고리즘의 흑인환자 위험 질환자 판정의 알고리즘 교정전(Original)과 교정후(Simulated)의 비교
 (Fig. 5) Comparison between before correction (Original) and after correction (Simulated) in the risk diagnosis of disease among Black patients by medical algorithm

는 민간 의료 예측 알고리즘의 인종적 편향성을 지적한 바 있다. 대학병원의 2013년부터 2015년간의 환자 자료를 토대로 흑인 환자 6,079명, 백인 환자 43,539명의 의료기록을 의료 알고리즘에 반영하여 해당 알고리즘의 인종별 만성질환 진단 판정 내용을 비교하였다. 알고리즘에 대한 비교결과 흑인 환자는 동일한 질환의 백인 환자보다 상대적으로 건강한 판정을 받을 가능성이 높은 것으로 나타났다. 이는 해당 알고리즘이 질병이 아닌 의료비 지출을 기준으로 활용하기 때문으로 흑인환자에게는 적은 의료비가 지출되면서 동일한 질환의 백인 환자보다 건강하다는 결론을 내리는 것이다. 한편 해당 알고리즘에 대한 편향성을 교정한 결과 17.7%에서 46.5%까지 의료비 지원을 받을 수 있는 흑인 환자의 비율이 증가하는 것으로 나타났다. 연구자들은 이러한 편향을 제거하기 위해 투입되는 데이터를 교체하고, 건강 이슈에 대해 총체적이고 다차원적인 관점을 반영할 필요가 있음을 강조한다.

V. 사례분석의 시사점

데이터 불평등의 세 가지 유형에 따른 사례에 대한 시사점은 다음과 같다.

첫째, 데이터 이용능력에 따른 불평등 해소를 위해 인프라 보급 확대 및 공공기관의 데이터 인식 제고가 요구된다. 사회적·경제적 불평등으로 인해 저개발국 및 소외계층이 특히 데이터 불평등의 영향을 받을 수 있음을 확인하였다(Naude & Vinuesa, 2021; PARIS 21 & MIF, 2021; Worldbank, 2021). 디지털 전환으로 인한 데이터 가치의 증대에도 불구하고, 활용할 수 있는 데이터가 없거나 품질이 떨어져 데이터를 활용할 수 없는 상황이 기존의 격차를 더욱 고착화 및 악화시키는 것으로 확인된다. 그리고 데이터 이용능력이 충족되지 않는 상황에서 데이터 활용은 해당 국가 및 집단 내에서 새로운 격차를 야기할 수 있음을 유의해야 한다. COVID-19 당시 감비아의 사례와 같이 스마트폰이 없는 저소득층의 데이터는 누락되기 때문에 정책결정자

의 의사결정과정에서 배제되고 잘못된 산출물로 이어질 수 있다(Worldbank, 2021). 한편 국가 간 정부의 데이터 이용능력 외에도 국가 내 지방정부간의 데이터 이용능력도 인프라 및 지식의 수준에 따라 차이가 존재한다(Seo, 2023a). 데이터 기반 행정은 데이터 이용능력에 따라 다양한 차원에서 불평등을 야기할 수 있다. 데이터 이용능력 격차를 줄이기 위한 관련 인프라의 보급과 공공기관의 데이터에 대한 인식 제고와 교육이 요구된다.

둘째, 데이터 관리에 따른 불평등 해소를 위해 데이터 제공자의 권리를 강화해야 한다. 디지털 전환으로 인해 데이터 기반 사회가 등장하면서, 데이터를 위시한 새로운 권력계층이 부상함으로써 인한 데이터 불평등 양상을 확인하였다(Cinnamon, 2020; Hartzog & Selinger, 2015; Kim, 2020; McCarthy, 2016; Money today, 2023; News1, 2019). 공공부분의 경우 기존에도 일반시민에 대한 정보를 합법적으로 수집 및 관리해 왔다면, 오늘날 디지털 기술의 발전으로 더 많은 정보를 수집하고, 이를 통해 시민의 민감한 부분도 파악이 가능해졌다. 중국의 사회신용시스템과 COVID-19 당시 각국 정부의 데이터 활용 현황은 공공기관과 시민의 데이터에 따른 불균형관 관계를 보여주고 있다(Liang, et al., 2018; Seo, 2021). 민간부문은 데이터 기반 사회에서 특히 수혜를 받는 상황인데, 다양한 제품 및 서비스들이 디지털 환경을 근간으로 이용자들이 의도적 혹은 비의도적으로 제공하는 방대한 데이터를 기업들이 활용하면서 가치를 창출하고 있기 때문이다. 민간기업에 소비자가 제공하게 되는 데이터는 특히 개인의 선호 및 취향, 사고 등을 반영하는 경우가 많기 때문에, 공공기관보다 개인의 특징을 더욱 잘 파악할 수 있다. 한편 데이터 관리 차원의 불평등은 기존의 사회·경제적 격차 등 소외계층 외에도 디지털 환경의 보편화로 소외계층이 아닌 사람들도 구조적으로 피해를 입을 수 있어 보다 광범위하게 발생될 수 있다. 데이터 관리에 따른 불평등을 해소하기 위한 법률 및 독립기구 등 제도적 방안이 제시된다(Austin & Lie, 2021; Lee, 2020). 우선적으로 데이

터에 대한 권리는 그것을 수집 관리한 측이 아닌 데이터를 제공하는 측에 있다는 개인정보자기결정권(자기정보통제권) 개념의 확대가 요구된다(Mehr, 2017). 개인정보자기결정권에 대한 인식이 공고해야 데이터의 오남용 등에 따른 문제 인식과 각종 제도적 수단을 활용할 수 있기 때문이다.

셋째, 데이터 활용에 따른 불평등 해소를 위해 데이터에 대한 과의존성을 줄이고, 다양한 차원의 데이터 확보 노력이 필요하다. 관련 사례를 통해 데이터 이용 능력과 연관되는 데이터의 누락, 편향 등의 문제로 인해 도출된 산출물(제품, 서비스, 정책 등)이 야기하는 새로운 격차를 확인하였다(Ada Lovelace Institute, 2021; BBC, 2021; Davaki, 2023; Obermeyer, et al., 2019). 데이터 활용에 따른 불평등은 현재 데이터 이용 능력에 있어 선도적인 국가에서 주로 확인되는데, 데이터 기반 사회를 구축하기 위해서는 국가의 디지털 역량에 의존하기 때문이다. 데이터 활용에 따른 불평등의 주요 문제는 데이터 활용 주체가 자신이 활용하는 데이터가 충분하고 오류가 없으며 이를 활용한 결과물도 바람직하다는 착각을 가질 수 있다는 것이다. 이는 데이터와 같은 증거기반 의사결정체계가 기존의 의견 기반의 사결정보다 객관적이고 과학적이라는 인식에서 비롯되는데(Davies, 2004; Sutcliffe & Court, 2005), 문제는 데이터 및 분석도구 역시 객관성을 담보하기 어렵다. 데이터는 특정 현상을 대표하는 요소로 활용되지만 그러한 현상 자체가 이미 다양한 요인에 의해 이미 편향될 수 있음을 지각하는 것은 쉽지 않다. 앞서 미국 의료 예측 알고리즘, AMAZON 구직자 이력서 분석 시스템 등이 야기한 차별 문제는 데이터도 객관적일 수 없음을 보여주며, 기존에 고착화된 사회의 불평등 및 격차를 반영하고 있음을 알 수 있다. 따라서 데이터 활용 주체의 데이터 기반 체계에 대한 과도한 의존성을 경계하고, 다양한 관점에서 포괄적인 데이터 활용 능력이 요구된다(Obermeyer, et al., 2019). 데이터 역시 사회 내 이해관계자간 상호작용의 결과물이고 이는 사회적으로 우월한 위치에 있는 계층에게 상대적으로 유리하게 구성

될 수 있음을 인지해야 한다. 따라서 활용하는 데이터에 대한 교차검증은 물론, 새로운 시각에서 현상을 보여주는 새로운 데이터를 확보할 필요가 있다. 결과적으로 데이터 기반 사회를 구축하려는 정부의 데이터에 대한 균형적 시각이 요구된다.

VI. 결론 및 한계

오늘날 디지털 전환의 확산에 따라 사회 전 분야가 데이터 기반 사회로 변모하면서, 사회 전 분야의 산출물 생산은 물론 소통과 참여 등도 데이터에 근거하고 있다. 기존의 정보격차와 마찬가지로 데이터에 의한 새로운 불평등에 대응해야 하지만 현재까지 데이터 불평등에 대한 논의는 디지털 기기 접근성과 같은 정보격차와 비교하여 부족한 실정이다. 이에 본 연구는 데이터 불평등에 대한 개념 및 현황을 확인하고자 문헌 및 사례를 참조하고, 데이터 불평등을 세 가지 유형으로 구분하여 제시하였다. 이에 산재 되어 있는 데이터를 수집하고 분석할 수 있는 능력인 데이터 이용능력에 의해 야기되는 불평등, 데이터 제공자로부터 데이터 수집 및 활용 주체의 데이터 운영을 의미하는 데이터 관리에 의해 야기되는 불평등, 데이터를 통한 산출물 등을 창출하는 과정인 데이터 활용에 의해 야기되는 불평등에 따른 데이터 불평등 사례를 제시하였다. 데이터 불평등은 사회경제적 격차로 촉발되는 정보격차와 상당부분 중첩면서도 데이터 기반 사회가 잉태한 새로운 데이터 권력계층에 의한 불균형한 관계와 같이 다양한 형태의 불평등을 보여주고 있다. 특히 디지털 환경이 전 세계적으로 보편화되면서, 데이터에 대한 의존성이 심화되고 있기 때문에, 데이터 기반 사회에 대한 균형적인 시각이 필요하다.

본 연구는 현재 데이터 불평등에 대한 학술적 개념이 현재까지 체계화되지 않은 상황에서 관련 사례를 중심으로 탐색적 연구를 수행하였다. 데이터 불평등에 대한 용어 및 정의도 연구자에 따라 관점이 차이가 나는 상황에서 가능한 데이터로 인해 야기되는 다양한 불평등을 다루고자 했으나, 일부 반영이 미흡한 부분이 있을 수

있음을 밝힌다. 그리고 데이터 불평등을 보여줄 객관적인 자료가 기존 정보격차와 비교하여 상대적으로 부족하다는 것이다. 일부 데이터 인프라 관련 현황을 제외하고는 객관적인 통계를 근거로 한 것이 아닌 사건을 중심으로 다루었다는 한계점을 가지고 있다. 이는 데이터 불평등이 기존의 정보격차를 판단하는 기기에 대한 보유 및 접근만으로는 설명할 수 없기 때문에 불가피한 측면이 있다. 또한 데이터 불평등이 발생하는 것이 상당수가 저개발국 및 소외계층과 관련되기 때문에 관련 내용을 확인할 수 있는 데이터가 부족한 것이 사실이다. 탐색적 연구로서 본 연구가 지닌 한계점이 있으나, 국내의 경우 스마트폰 등 디지털 기기에 대한 접근성과 비교하여 현재까지 데이터 불평등에 대한 연구가 미진한 상황이기 때문에, 관련 연구의 필요성을 제기한 것에 의의를 두고자 한다.

■ References

- Ada Lovelace Institute (2021). The data divide.
- Anant, V., Donchak, L., Kaplan, J. & Soller, H. (2020). "The consumer-data opportunity and the privacy imperative." Retrieved from McKinsey website: <https://www.mckinsey.com/business-functions/risk-and-resilience/our-insights/the-consumer-data-opportunity-and-the-privacy-imperative>.
- Austin, L. & Lie, D. (2021). "Data trusts and the governance of smart environments: Lessons from the failure of sidewalk labs' urban data trust." *Surveillance & Society*, 19(2), 255-261.
- BBC. (2020). "UK passport photo checker shows bias against dark-skinned women." <https://www.bbc.com/news/technology-54349538> (Retrieved on October 8).
- BBC. (2021). "Facebook apology as AI labels black men 'primates'." <https://www.bbc.com/news/technology-58462511> (Retrieved September 6).
- Belanger, F. & Carter, L. (2009). "The impact of the digital divide on e-government use." *Communications of the ACM*, 52(4), 132-135.
- Britz, J. J. (2004). "To know or not to know: A moral reflection on information poverty." *Journal of Information Science*, 30(3), 192-204.
- Brownlow, J., Zaki, M., Neely, A. & Urmetzer, F. (2015). "Data and analytics-data-driven business models: A Blueprint for Innovation." *Cambridge Service Alliance*, 7(February), 1-17.
- Bulger, M., Taylor, G. & Schroeder, R. (2014). *Data-driven business models: challenges and opportunities of big data*. Oxford Internet Institute. Research Councils UK: NEMODE, New Economic Models in the Digital Economy. <https://www.oii.ox.ac.uk/news-events/reports/data-driven-business-models-challenges-and-opportunities-of-big-data/>
- Cinnamon, J. (2020). "Data inequalities and why they matter for development." *Information Technology for Development*, 26(2), 214-233.
- Pentland, A. S. (2013). "The data-driven society." *Scientific American*, 309(4), 78-83.
- Davaki, K.(2023). *The impact of the gender data gap on consumer protection*. European Parliament report.
- Davies, P. (2004). Is evidence-based government possible? Jerry Lee Lecture, presented at the 4th Annual Campbell Collaboration Colloquium, Washington DC
- Fitzgerald, M., Kruschwitz, N., Bonnet, D. & Welch, M. (2014). "Embracing digital technology: A new strategic imperative." *MIT sloan management review*, 55(2), 1.
- Frankiewicz, B. & Chamorro-Premuzic, T. (2020). "Digital transformation is about talent, not technology." *Harvard Business Review*, 6(3), 1-6.
- Gartner (2016). *Top 10 Strategic Technology Trends for 2017*.
- Giest, S. & Samuels, A. (2020). "'For good measure': data gaps in a big data world." *Policy Sciences*, 53(3), 559-569.
- Ginsberg, W. (2011). *The Obama Administration's Open Government Initiative*. Issues for Congress, CRS

- Report for Congress Prepared for Members and Committees of Congress. In Library of Congress. Congressional Research Service.
- Hartmann, P., Zaki, M., Feldmann, N. & Neely, A. (2014). Big data for big business? A taxonomy of data-driven business models used by start-up firms. http://cambridgeservicealliance.blogspot.co.uk/2014/04/big-data-for-bigbusiness_3.html.
- Hartzog, W. & Selinger, E. (2015). Surveillance as loss of obscurity. *Wash. & Lee L. Rev.*, 72, 1343.
- Hotz, N. (2022). Why Big Data Science & Data Analytics Projects Fail. <https://www.datascience-pm.com/project-failures/>
- Ismail, M. H., Khater, M. & Zaki, M. (2017). "Digital business transformation and strategy: What do we know so far." *Cambridge Service Alliance*, 10(1), 1-35.
- Janssen, M., Charalabidis, Y. & Zuiderwijk, A. (2012). "Benefits, adoption barriers and myths of open data and open government." *Information Systems Management*, 29(4): 258-268.
- Joint Ministries. (2020). *Korea New Deal: National Strategy*
{관계부처 합동 (2020). <「한국판 뉴딜」 종합계획>}
- Katz, J. & Aspden, P. (1997). "Motivations for and barriers to Internet usage: Results of a national public opinion survey." *Internet Research*, 7(3), 170-188.
- Kim, J. (2017). George Orwell's Nightmare: *China's Social Credit System*, Seoul: Asan Institute For Policy Studies.
{김진우 (2017). <조지 오웰의 악몽: 중국의 사회신용시스템>. 서울: 아산정책연구원.}
- Kim, M. (2020). "Data Divide in the Platform Data Ecosystem: Beyond digital inequality." *Communication Theories*, 16(4), 5-45.
{김미경 (2020). 플랫폼 데이터 생태계에서 데이터 격차: 디지털 불평등을 넘어. <커뮤니케이션 이론>, 16권 4호, 5-45.}
- Lee, D. (2020). *The Importance of Social Consensus through Toronto Smart city cases*. Daegu: National Information Society Agency.
- {이대철 (2020). <토론토 스마트시티 사례로 보는 스마트시티 데이터 활용에 대한 사회적 합의의 중요성>. 대구: 한국지능정보사회진흥원.}
- Lee, S. (2019). "Chinese and Korean University Students' Perceptions of China's Social Credit System: Everyday Experience of Datafication." *The Journal of Asian Studies*, 22(4), 225-248.
- {이승은 (2019). 중국 사회신용시스템에 대한 한중 대학생 인식 비교: 데이터화의 일상경험을 중심으로. <아시아연구>, 22권 4호, 225-248.}
- Liang, F., Das, V., Kostyuk, N. & Hussain, M. M. (2018). "Constructing a data-driven society: China's social credit system as a state surveillance infrastructure." *Policy & Internet*, 10(4), 415-453.
- Lucas, P., J. & Robinson, R., Treacy, L. (2020). *What Is Data Poverty?*
- Lutz, C. (2019). "Digital inequalities in the age of artificial intelligence and big data." *Human Behavior and Emerging Technologies*, 1(2), 141-148.
- McCarthy, M. T. (2016). "The big data divide and its consequences." *Sociology Compass*, 10(12), 1131-1140.
- McIntosh, D. (2018). "We need to talk about data: how digital monopolies arise and why they have power and influence." *J. Tech. L. & Pol'y*, 23, 185.
- Mehr, H. (2017). *Artificial Intelligence for Citizen Services and Government*. Harvard Ash Center.
- Milan, S. & Trere, E. (2020). "The Rise of the Data Poor: The COVID-19 Pandemic Seen From the Margins." *Social Media+ Society*, 6(3), 2056305120948233.
- Ministry of Science and ICT (2024). *Wireless communication service statistics status (December 2023)*.
- {과학기술정보통신부. (2024). <무선 통신서비스 통계 현황> (2023년 12월 기준).}
- Moneytoday (2023). "'Beep' disaster text message is loud, but mom's phone is silent... 2G/3G phones that do not receive disaster notice." August 3.}
- {머니투데이 (2023). "삐익" 재난문자 요란한데 엄마 폰은

- 잠잠...알림 못 받는 2G·3G폰, 8월 3일.}
- Naude, W. & Vinuesa, R. (2021). "Data deprivations, data gaps and digital divides: Lessons from the COVID-19 pandemic." *Big Data & Society*, 8(2), 205395172111025545.
- Neff, G. & Nagy, P. (2016). "Talking to bots: Symbiotic agency and the case of Tay." *International Journal of Communication*, 10, 4915-4931.
- New, J., Castro, D. & Beckwith, M. (2017). *How national governments can help smart cities succeed*. Washington, DC, USA: Center for Data Innovation.
- News1 (2019). "The disaster emergency safety evacuation text message does not work on 2G phones in Samcheok City." October 3.
- {News1 (2019). "삼척시 재난안전대피 문자 2G폰엔 먹통." 10월 3일. <https://www.news1.kr/articles/?3735355>}
- Norris, P. (2001). *Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide*. New York: Cambridge University Press
- Nuccio, M. & Guerzoni, M. (2019). "Big data: Hell or heaven? Digital platforms and market power in the data-driven economy." *Competition & Change*, 23(3), 312-328.
- Obermeyer, Z., Powers, B., Vogeli, C. & Mullainathan, S. (2019). "Dissecting racial bias in an algorithm used to manage the health of populations." *Science*, 366(6464), 447-453.
- O'Neil, C. (2016). *Weapons of math destruction. How big data increases inequality and threatens democracy*. Crown: New York.
- PARIS 21 & MIF (2021). BRIDGING THE DATA-POLICY GAP IN Africa. Working paper. https://paris21.org/sites/default/files/inline-files/Data-Policy%20Gap_Africa__FINAL_20210430.pdf
- Qi, X. (2020). "How next-generation information technologies tackled COVID-19 in China. In World Economic Forum." <https://www.weforum.org/agenda/2020/04/how-next-generation-information-technologies-tackled-covid-19-in-china>.
- Raposo, V. L. (2021). "Big Brother Knows that you are infected: wearable devices to track potential COVID-19 infections." *Law, Innovation and Technology*, 13(2), 422-438.
- Reuter, T. K. (2020). *Smart City Visions and Human Rights: Do They Go Together?*. Carr Center for Human Rights Policy Harvard Kennedy School.
- Reuters (2018). "Amazon scraps secret AI recruiting tool that showed bias against women." October 11, <https://www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKCN1MK08G>
- Richmond, B. (2006). *Introduction to Data Analysis Handbook*. Academy for Educational Development.
- Sawyer, S. (2008) "Data Wealth, Data Poverty, Science, and Cyber infrastructure." *Prometheus*, 26(4), 357-371.
- Serajuddin, U., Uematsu, H., Wieser, C., Yoshida, N. & Dabalen, A. (2015). *Data deprivation: another deprivation to end*. World Bank policy research working paper, (7252).
- Slynchuk, A. (2022). "Big Brother Brands Report: Which Companies Access Our Personal Data the Most?." <https://clario.co/blog/which-company-uses-most-data/>
- Smidt, H. J. & Jokonya, O. (2021). "The challenge of privacy and security when using technology to track people in times of COVID-19 pandemic." *Procedia Computer Science*, 181, 1018-1026.
- Schäfer, F., Gebauer, H., Gröger, C., Gassmann, O. & Wortmann, F. (2023). "Data-driven business and data privacy: Challenges and measures for product-based companies." *Business Horizons*, 66(4), 493-504.
- Seo, H. (2017). "A Critical Review on Open, Useful, Reusable Government Data Index by OECD with Level of Domestic Open Government Data: Focusing on Comparison with Open Data Barometer." *Informatization policy*, 24(2), 43-67.
- {서형준 (2017). 국내 공공데이터 개방수준을 통해서 본 OECD 의 Open, Useful, Reusable Government Data Index 에 대한 비판적 논의: Open Data

- Barometer와의 비교를 중심으로. <정보화정책>, 24권 2호, 43-67.}
- Seo, H. (2019). "A Preliminary Discussion on Policy Decision Making of AI in The Fourth Industrial Revolution." *Informatization policy*, 26(3), 3-35.
- {서형준 (2019). 4차 산업혁명시대 인공지능 정책의사결정에 대한 탐색적 논의. <정보화정책>, 26권 3호, 3-35.}
- Seo, H. (2021). "The Determinant Factors on Behavior of Information Security in Public Sector: Focusing on Mediating Effect of Perception of Information Security." *The Korean Journal of Public Administration*, 30(2), 173-207.
- {서형준 (2021). 공공부문 정보보안 행태에 미치는 영향요인: 정보보안 의식의 매개효과를 중심으로. <한국행정연구>, 30권 2호, 173-207.}
- Seo, H. (2023a). "Analysis of Municipal Ordinances for Smart Cities of Municipal Governments: Using Topic Modeling." *Informatization policy*, 30(1), 41-66.
- {서형준 (2023a). 지방자치단체의 스마트시티 조례 분석: 토픽모델링을 활용하여. <정보화정책>, 30권 1호, 41-66.}
- Seo, H. (2023b). "A Policy Study on the Implementation of Domestic Digital Platform Government: Focusing on the Classification of Domestic and Foreign Cases of Government as a Platform (Gaap)." *Informatization policy*, 30(4), 113-137.
- {서형준 (2023b). 국내 디지털플랫폼정부 구현을 위한 정책연구: 국내·외 플랫폼 정부 사례의 유형화를 중심으로. <정보화정책>, 30권 4호, 113-137.}
- Seo, H. & Myeong, S. (2016). "Research Trends and Issues on Digital Divide: Focusing on Korean Journals." *Journal of Korean Association for Regional Information Society*, 19(4), 151-187.
- {서형준·명승환 (2016). 국내 정보격차연구 동향: 국내 학술지를 중심으로. <한국지역정보학회지>, 19권 4호, 151-187.}
- Sorescu, A. (2017). "Data-driven business model innovation." *Journal of Product Innovation Management*, 34(5), 691-696.
- Sutcliffe, S. & Court, J. (2005). *Evidence-Based Policymaking: What is it? How does it work? What relevance for developing countries?* Overseas Development Institute.
- Tessian (2022). "30 Biggest GDPR Fines So Far (2020, 2021, 2022)." <https://www.tessian.com/blog/biggest-gdpr-fines-2020/>
- The Guardian (2016). "Microsoft 'deeply sorry' for racist and sexist tweets by AI chatbot." 26 March. <https://www.theguardian.com/technology/2016/mar/26/microsoft-deeply-sorry-for-offensive-tweets-by-ai-chatbot>
- The Seoul Shinmun Daily (2022). "Regional Severe Gap of Public WIFI...68 times difference between Seoul and Gangwon." October 4.
- {서울신문 (2022). "공공와이파이 지역 양극화 극심... 서울과 강원 68배 차이." 10월 4일.} https://www.seoul.co.kr/news/newsView.php?id=20221004500110&wlog_tag3=naver
- Thierer, A. (2000). "How free computers are filling the digital divide." *Heritage Foundation Backgrounder*, 1361, 1-21.
- Ubaldi, B., Gonzalez-Zapata, F. & Barbieri, M. P. (2020). *Digital Government Index: 2019 results*. Paris: OECD Publishing,
- van Ooijen, C., Ubaldi, B. & Welby, B. (2019). *A data-driven public sector*. OECD Working Papers on Public Governance, No 33.
- Vial, G. (2019). "Understanding digital transformation: A review and a research agenda." *The journal of strategic information systems*, 28(2), 118-144.
- Yannoukakou, A. & Araka, I. (2014). "Access to government information: Right to information and open government data synergy." *Procedia-Social and Behavioral Sciences*, 147, 332-340.
- Warren, M. (2007). "The digital vicious cycle: Links between social disadvantage and digital exclusion in rural areas." *Telecommunications Policy*, 31, 374-388.
- Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U. & Kortuem, G. (2016). "Creating an understanding of data literacy for a data-driven society." *The Journal of Community Informatics*, 12(3).
- Worldbank. (2021). *World development report 2021: Data for better lives*. Washington DC: The World Bank Group.